COMPLICATIONS IN CLINICAL TRIALS: BAYESIAN MODELS FOR REPEATED MEASURES AND SIMULATORS FOR NONADHERENCE

by

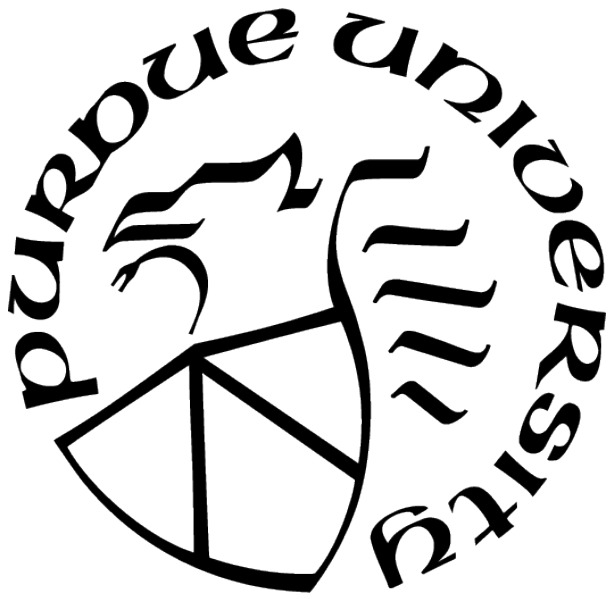
Ahmad Hakeem Bin Abdul Wahab

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Statistics

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Arman Sabbaghi, Chair

Department of Statistics

Dr. Marguerite O'Haire

Department of Comparative Pathobiology

Dr. Vinayak Rao

Department of Statistics

Dr. Bruce Craig

Department of Statistics

Approved by:

Dr. Jun Xie

ACKNOWLEDGMENTS

I thank my advisor, Dr. Arman Sabbaghi, whose amazing support, guidance, and mentorship helped shape this body of research. I thank the O'Haire Lab members (Dr. Marguerite O'Haire, Clare Jensen, and Leanne Nieforth), for providing much insight into such an exciting area of research and moral support with their amazing dogs and horses. I thank my research team (the "Armanys": Daniel Cardona, Yumin Zhang, Wenbin Zhu, Yueyun Zhang, Dominique Williams, Run Zhuang) for giving feedback on my research and presentations. I thank Dr. Stephen Ruberg and members of the Tripartite Team, who provided heated debates and discussions on causal models and their interpretations. I thank the department's amazing administrative staff for checking in on me and always having our best interests at heart. I thank my good friend Daniel Vasquez who stuck it through with me during our PhD years and provided mental support for me when times were rough. I thank Dr. Zachary Hass for his wonderful friendship and support, especially during trying times. I thank Dr. Tim Keaton who introduced me to board games and the many eating adventures. I thank the amazing Doug Crabill for helping me with all things computational and Cheryl Crabill for providing a nice respite from work.

To my off-campus family, I thank the Paddocks for letting me crash their place during Christmas and wonderful lunches with grandma and grandpa. I thank the Wells for hosting game nights and provided great company during Friday nights. I thank the brothers Brandon and Brody Gregg as well as their families for many game and Anime nights. I thank Auntie Sha and Uncle Simon who always had my back when things were getting sour.

To my blood family, I thank my parents Abdul Wahab Abdul Rahman and Kamariah Adnan for their unconditional love, who have always been encouraging me to finish my PhD and investing in my education. I thank my siblings (Nurhayati, Abdul Nasir, and Nuraini) for being there when I needed it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Clinical trials are the gold standard for inferring the causal effects of treatments or interventions. This thesis is concerned with the development of methodologies for two problems in modern clinical trials. First is analyzing binary repeated measures in clinical trials using models that reflect the complicated autocorrelation patterns in the data, so as to obtain high power when inferring treatment effects. Second is simulating realistic outcomes and subject nonadherence in Phase III pharmaceutical clinical trials under the Tripartite Framework.

Bayesian Models for Binary Repeated Data: The Bayesian General Logistic Autoregressive Model and the Polya-Gamma Logistic Autoregressive Model

Autoregressive processes in generalized linear mixed effects regression models are convenient for the analysis of clinical trials that have a moderate to large number of binary repeated measurements, collected across a fixed set of structured time points, for each subject. However, much of the existing literature and methods for autoregressive processes on repeated binary measurements permit only one order and only one autoregressive process in the model. This limits the flexibility of the resulting generalized linear mixed effects regression model to fully capture the dynamics in the data, which can result in decreased power for testing treatment effects. Nested autoregressive structures enable more holistic modeling of clinical trials that can lead to increased power for testing effects. We introduce the Bayesian General Logistic Autoregressive Model (BGLAM) for the analysis of repeated binary measures in clinical trials. This model extends previous Bayesian models for binary repeated measures by accommodating flexible and nested autoregressive processes with non-informative priors. We describe methods for selecting the order of the autoregressive process in BGLAM based on the Deviance Information Criterion (DIC) and marginal log-likelihood, and develop an importance sampling-weighted posterior predictive p-value to test for treatment effects in BGLAM. The frequentist properties of BGLAM compared to existing likelihood- and non-likelihood-based statistical models are evaluated by means of extensive simulation studies involving different data generation mechanisms. We apply our model for data collected from a clinical trial on the effects of Service Dogs for reducing PTSD symptoms of United States Veterans. Ultimately, on the basis of simulation studies and the

real-life case study, we conclude that BGLAM provides a more effective and comprehensive approach for testing treatment effects in clinical trials with repeated binary measures and complex autoregressive patterns.

Two features of BGLAM that can limit its practical application are the computational effort involved in executing it and the inability to integrate added heterogeneity across time in its autoregressive processes. We develop the Polya-Gamma Logistic Autoregressive Model (PGLAM) for addressing these limiting features. This new model enables the integration of additional layers of variability through random effects and heterogeneity across time in nested autoregressive processes. Furthermore, PGLAM is computationally more efficient than BGLAM because it eliminates the need to use the complex types of samplers for truncated latent variables that is involved in the Markov Chain Monte Carlo algorithm for BGLAM. We exhibit via additional, extensive simulation studies that the new features introduced by PGLAM do not adversely affect its frequentist properties in a significant manner. Furthermore, we demonstrate that PGLAM better captures complex layers of variability compared to existing likelihood-based models, both in terms of yielding better power for testing treatment effects and higher coverage for the confidence intervals for the treatment effects

CITIES: Clinical Trials with Intercurrent Events Simulator

Clinical trials are the gold standard for evaluating the efficacy of new pharmaceutical interventions. Although clinical trials are designed with strict controls, inevitably complications will arise during the course of the trials. One significant type of complication is missing subject outcomes due to subject drop-out or nonadherence during the trial, which are referred to in general as *intercurrent events*. This complication can arise from, among other causes, adverse reactions, lack of efficacy of the assigned treatment, administrative reasons, and excess efficacy from the assigned treatment. Intercurrent events typically confound causal inferences on the effects of the treatments under investigation because the resulting missingness that occurs corresponds to a Missing Not at Random missing data mechanism. The missingness is driven based on latent strata of patients characterized by their adherence behaviors under the different possible assigned treatments. These latent strata must be taken into account in order to obtain valid causal inferences on the causal effects of the

receipt of treatment, and not merely the *assignment* of treatment. The latter type of effect is typically considered in standard intention-to-treat (ITT) analyses, and has several flaws that are described in the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9(R1) amendment to the original ICH E9 guidelines for clinical trials. As a consequence of the amendment to ICH E9, the pharmaceutical industry is increasingly focused on developing methods for obtaining valid causal inferences on the receipt of treatment in clinical trials with intercurrent events. However, it is extremely difficult to compare the frequentist properties and performance of these competing methods, as real-life clinical trial data cannot be easily accessed or shared, and as the different methods consider distinct assumptions for the underlying data generating mechanism in the clinical trial. We develop a novel simulation model for clinical trials with intercurrent events. Our simulator operates under the Rubin Causal Model. We implement the simulator by means of an R Shiny application. This app enables users to control patient compliance through different sources of discontinuity with varying functional trends, and understand the frequentist properties of treatment effect estimators obtained by different models for various estimands. Under our simulation, the treatment effect accounts for intercurrent events in clinical trials with multiple endpoints. Based on the application of our simulator to capture data from two real-life clinical trials, we conclude that our new data generating mechanism is a convenient tool for practitioners in the pharmaceutical industry to compare the methods they develop for analyzing clinical trials on the same, comprehensive setting.

1. BGLAM: A BAYESIAN GENERAL LOGISTIC AUTOREGRESSIVE MODEL FOR CORRELATED BINARY OUTCOMES

1.1 Introduction

Binary repeated measures studies are experiments or observational studies in which each subject has several binary values for an outcome variable observed across time. These studies are prevalent across a wide variety of domains, ranging from medicine and biology to the social sciences [1]. One significant challenge in modeling binary repeated measures is the incorporation of an appropriate correlation structure for the outcomes. Failure to account for correlation in the repeated measure outcomes generally results in incorrect standard errors, which can yield incorrect Type I error rates and low power when performing statistical inferences [2]. In practice, this challenge is addressed by incorporating correlation structures into a logistic regression model. This is due to the advantage of these models providing interpretable regression coefficients, which is recognized among statisticians as well as subject-matter specialists [3], [4].

Several frequentist statistical methods for binary repeated measures studies exist that yield marginal interpretations similar to logistic regression while accounting for the correlations between the binary outcomes. Chief among them are the methods of generalized estimating equations [GEE; 5]–[8] and generalized linear mixed models [GLMM; 9]. GEE methods can enable direct inferences on marginal logistic models for binary repeated measures, and are robust to potential misspecifications of the correlation structure. However, a disadvantage of GEE methods is that a large number of observations are generally necessary for the application of their asymptotics-based inferences, with small data sets yielding inaccurate inferences [2], [10, p. 170]. In contrast to GEE methods, GLMM methods typically involve conditional models that utilize estimates and inferences based on a likelihood function, which yields several practical advantages. Examples of such advantages are that likelihood-based tests are well-defined, valid assessments of model fit can be performed, and model selection techniques based on the likelihood can be implemented [2, p. 144]. A poten-

tial disadvantage of existing GLMM methods is their reliance on likelihood approximations [11], [12], e.g., Gauss-Hermite quadrature [13]. In addition, GLMM methods that involve marginal models utilize pseudolikelihood functions, and these functions do not correspond to a true likelihood function. The corresponding disadvantage is that these methods can yield biased covariance estimators, especially for binary data [14, p. 2219]. Finally, a common disadvantage for the popular implementations GLMM methods in popular programming environments such as SAS® PROC GLIMMIX, respectively, is that they are limited to modeling only Autoregressive(1) or AR(1) structures.

One statistical method that addresses all of the limitations associated with existing GEE and GLMM methods is the Bayesian Multivariate Logistic Regression [BMLR; 4] model. This model uses a latent multivariate distribution for the repeated binary outcomes that captures the correlation structure of interest, with the underlying parameters being interpretable on the log odds scale (formal details on the implementation are in Section 1.3). Large samples of data are not necessary to justify the Bayesian inferences obtained from the BMLR model. Instead, direct and straightforward uncertainty assessments for inferences based on the posterior distributions of its model parameters are obtained without the need to appeal to asymptotic covariance matrices. Also, as the standard specification of the prior distribution for this model given by O'Brien and Dunson [4] enables the computation via Markov Chain Monte Carlo (MCMC) methods of a corresponding proper posterior distribution. A distinct advantage of the Bayesian paradigm for the BMLR model compared to existing frequentist methods is that it enables practitioners to integrate substantive prior information from domain experts into the analysis of repeated measures studies.

Despite its advantages, the standard BMLR model typically fails to recover AR correlation structures for a large number of repeated measures (Figure 1.1). The consequence of this issue is that standard errors and Type I error rates for inferences on logistic regression coefficients will be incorrect. To illustrate this failure, we consider a simulated data set with 100 subjects generated according to an AR(1) process via the t-copula inherent in BMLR, with autocorrelation parameter $\rho = 0.5$ (formal details on this data generating mechanism are in Section 1.3). The standard BMLR model accurately infers the autocorrelation structure for five repeated measures. However, in the cases of 15 or 30 repeated measures the recovered

autocorrelation matrix does not correspond to the true AR(1) autocorrelation structure. Indeed, for the case of 15 repeated measures, the inferred correlations lose any semblance to the underlying structure of the specified autoregressive structure. Such striking discrepancies are especially concerning for the validity of the analyses of the growing number of studies that involve a large number of repeated measures, such as those frequently conducted in the fields of aging research [15], labor market surveys [16], geographical surveys [17], and metabolomics [18].

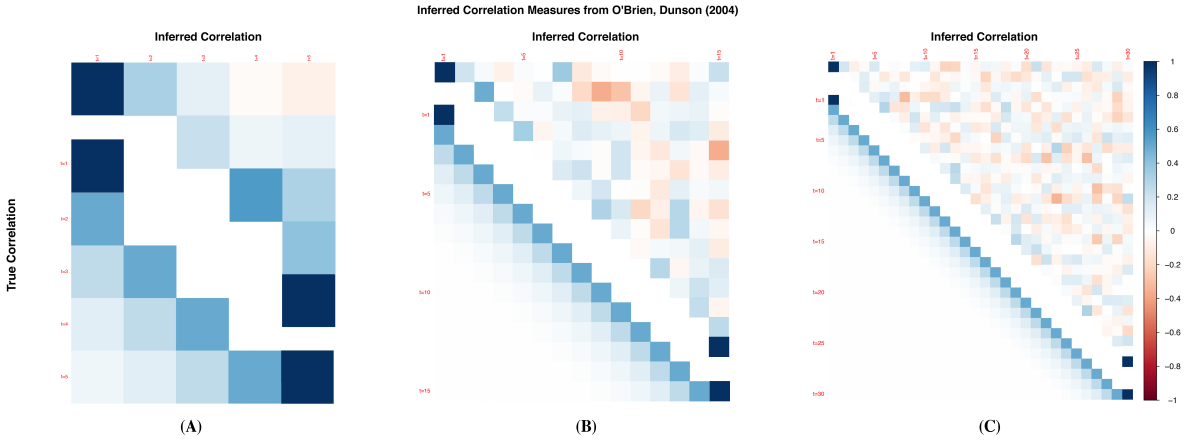


Figure 1.1. Comparisons of the true and inferred correlation structures for data generated using the data generating mechanism inherent in the BMLR model via t-copulas. In each sub-figure, the cells below the diagonal are the true correlation values between two time points, and those above the diagonal are the inferred correlation values based on the fitted BMLR model. The empty diagonal is meant to separate the true and inferred correlation values. Each data set consists of 100 subjects, with the number of repeated time points set 5 for Figure 1.1(a), 15 for Figure 1.1(b), and 30 for Figure 1.1(c). The fitted BMLR model fails to recover the true correlation structure as the number of repeated measures increases.

Prior work has been done on extending the original BMLR model. Nooraee, Abegaz, Ormel, *et al.* [19] expanded on the BMLR model and demonstrated that their expanded model outperformed the method of GEE on ordinal data. However, their results were only demonstrated for data with up to three time points, and they did not demonstrate that they would be able to recover the true autoregressive structure for a large number of repeated measurements without additional restrictions or assumptions. Hirk, Hornik, Vana, *et al.* [20]

extended the different classes of autoregressive structures beyond those that were considered in [4], but only up to an AR(1) setting. Paul, Maity, and Maiti [21] fitted the BMLR model on time series data with varying autoregressive settings, but the parameterization of the correlation and effects in this particular model differ from those that are involved in the analyses of repeated measures data. Furthermore, the rigidity of such time series models permits only one autoregressive structure for a given series or subject.

We directly address the issues inherent with the standard BMLR model by specifying a distinct prior distribution for the autoregressive temporal structures of the binary outcomes. Specifically, we place a prior on the covariance matrix that is sufficiently flexible for accommodating a wide range of autoregressive process. An advantage of our extension of the standard BMLR model is that we are better able to model correlated binary outcomes with a large number of repeated measurements that may have nested autoregressive structures of arbitrary order, and account for correlations between the nested autoregressive structures as well. We refer to our new model as the Bayesian General Logistic Autoregressive Model (BGLAM).

We commence in Section 1.2 by providing the formal notation and assumptions we consider for binary repeated measures studies, and a detailed review of existing GEE and GLMM models. Our BGLAM model is formally described in Section 1.3. Simulation studies that compare the performance of the BGLAM model to those of GEE, GLMM, and the standard BMLR methods for AR(1) and AR(2) processes are described in Section 1.4. An real-life application of the BGLAM model is provided in Section 1.5. In this case study, the BGLAM model is fitted to data from a National Institutes of Health (NIH)-sponsored study on the effects of service dogs on the daily lives of post-9/11 military veterans diagnosed with post-traumatic stress disorder (PTSD). Concluding remarks on binary repeated measures studies, the BGLAM model, and potential extensions for future investigation are provided in Section 1.6.

1.2 Background

1.2.1 Notation

Let $\boldsymbol{y}_{\mathrm{i}} = (y_{\mathrm{i}1}, \ldots, y_{\mathrm{i}n_{\mathrm{i}}})$, $\mathrm{i} = 1, \ldots, N$ where $y_{\mathrm{i}n_{\mathrm{i}}}$ is the n_{i}-th observation for subject i. There are n_{i} repeat measurements for subject i and a total of $\sum_{\mathrm{i}=1}^{N} n_{\mathrm{i}}$ measurements with the associated matrix of baseline covariates $\boldsymbol{x}_{\mathrm{i}} = (\boldsymbol{x}_{\mathrm{i}1}, \ldots, \boldsymbol{x}_{\mathrm{i}p})$. This is motivated by the clinical trial data to which we apply this model in Section 1.5, as covariates were collected at baseline and we are more interested in their average effect across time throughout the study.

For an individual i with n_{i} repeat measures, the $(\mathrm{j}, \mathrm{j}+k)$th element of the correlation matrix $\underset{n_{\mathrm{i}} \times n_{\mathrm{i}}}{\boldsymbol{R}_{\mathrm{i}}}$ is defined as the marginal correlation $r_{\mathrm{j},\mathrm{j}+k} = \mathrm{Cor}(y_{\mathrm{j}}, y_{\mathrm{j}+k})$, $\mathrm{j}, k = 1, \ldots, n_{\mathrm{i}} - 1$. Alternatively, $\boldsymbol{R}_{\mathrm{i}}$ can also be parameterized using partial autocorrelations, defined as $\rho_{\mathrm{j},\mathrm{j}+k} = \mathrm{Cor}(y_{\mathrm{j}}, y_{(\mathrm{j}+k)} | y_l, \mathrm{j} < l < \mathrm{j}+k)$, $\mathrm{j}, k = 1, \ldots, n_{\mathrm{i}}-1$. In other words, the partial autocorrelation between time points j and $\mathrm{j}+k$ is the pure and unconfounded correlation between these two points after removing for the effects of all intermediate time points while the marginal correlation is the final realized correlation between j and k, including those from intermediate time points. A brief exposition on partial autocorrelations will be provided in Section 1.3.1.

1.2.2 Literature Review

A prominent method for modelling binary repeat measurements uses the Generalized Linear Mixed Model [9] using the GLIMMIX procedure in SAS®. There are two ways to model repeated measures in GLIMMIX: the G-side and the R-side [2]. A G-side or conditional model would yield $g\Big(\mathbb{E}[\boldsymbol{Y}|\boldsymbol{\gamma}]\Big) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma}$, where g is a differentiable monotonic link function chosen based on the assumed distribution of the response $\boldsymbol{Y}$, $\boldsymbol{\beta}$ is the vector of fixed effects for its associated design matrix $\boldsymbol{X}$ and $\boldsymbol{\gamma}$ is the vector of random effects for its associated design matrix $\boldsymbol{Z}$ that contains the the different levels for the random factor time. We assume $\boldsymbol{\gamma} \sim N(\boldsymbol{0}, \boldsymbol{G})$, where $\boldsymbol{G}$ would be the user-specified covariance structure for the repeat measures such as autoregressive(1).

In contrast, the R-side or marginal model is the same as the G-side, just that there is no longer the random effect $\boldsymbol{\gamma}$. In addition, we assume the following about the marginal

variance $\text{Var}[\boldsymbol{Y}] = \boldsymbol{A}^{\frac{1}{2}}\boldsymbol{R}\boldsymbol{A}^{\frac{1}{2}}$, where $\boldsymbol{R}$, also known as the working correlation matrix, would be the user-specified covariance structure for the repeat measures and $\boldsymbol{A}$ is a diagonal matrix containing the variance function, defined as the variance of a random variable as a function of its mean [22]. We can see that the G-side models the conditional expectation $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{\gamma}]$ while the R-side models the marginal expectation $\mathbb{E}[\boldsymbol{Y}]$. Hence, the G-side is oft called the conditional model while the R-side is called the marginal model [14].

A lucrative feature of conditional models is that their estimates are based on the true likelihood via Integral Approximations [11], [12] such as Laplace Approximation [23] and Gauss-Hermite Quadrature [13] while those of the marginal models are based on pseudo likelihoods [14]. As a direct consequence, this also means that marginal models are more robust to model misspecifications than conditional models.

The GEE of Liang and Zeger [5] used to estimate the regression parameters $\boldsymbol{\beta}$ for correlated data is given by solving the score function $\boldsymbol{S}(\boldsymbol{\beta}) = \sum_{\text{i}=1}^{N} \boldsymbol{D}_{\text{i}}^{T}\boldsymbol{V}_{\text{i}}^{-1}(\boldsymbol{y}_{\text{i}} - g^{-1}(\boldsymbol{x}_{\text{i}}\boldsymbol{\beta})) \underset{set}{=} 0$, where $\boldsymbol{D}_{\text{i}} = \frac{\partial}{\partial\boldsymbol{\beta}} g^{-1}(\boldsymbol{x}_{\text{i}}\boldsymbol{\beta})$, $\boldsymbol{V}_{\text{i}}$ is the covariance matrix of $\boldsymbol{y}_{\text{i}}$, $\boldsymbol{x}_{\text{i}}$ is the design matrix across all time points for subject i and g is a differentiable monotonic link function chosen based on the assumed distribution of the response $\boldsymbol{y}_{\text{i}}$. GEE estimates the $\boldsymbol{\beta}$ parameters by maximizing the log likelihood function over all subjects $\boldsymbol{L}$ with respect to the regression parameters using a ridge-stabilized Newton-Raphson algorithm [24].

The working correlation matrix $\boldsymbol{R}_{\text{i}}$ here is user specified, yielding the following covariance matrix $\boldsymbol{V}_{\text{i}} = \nu\boldsymbol{A}_{\text{i}}^{\frac{1}{2}}\boldsymbol{W}_{\text{i}}^{-\frac{1}{2}}\boldsymbol{R}_{\text{i}}\boldsymbol{W}_{\text{i}}^{-\frac{1}{2}}\boldsymbol{A}_{\text{i}}^{\frac{1}{2}}$, where ν is the overdispersion parameter, $\boldsymbol{A}_{\text{i}}$ is the variance function and $\boldsymbol{W}_{\text{i}}$ is a user-specified weight matrix that defaults to an identity matrix $\boldsymbol{I}$. GEEs integrate out the random effects, thus they are often used to model population averaged effects. Note that the marginal variance of $\boldsymbol{y}_{\text{i}}$ in GEEs is very similar to that of the R-side, assuming $\phi = 1$ and $\boldsymbol{W}_{\text{i}} = \boldsymbol{I}$. Thus, results for R-side models in GLMM should be similar, if not the same, to GEE results.

Alternatively, the multivariate logistic distribution [4] uses a multivariate T to account for the correlation while still maintaining logistic marginals. For a given individual i, the probability distribution function is given by

$$\mathscr{L}_{n_{\mathrm{i}},v}(\boldsymbol{z}_{\mathrm{i}}|\boldsymbol{\mu}_{\mathrm{i}},\boldsymbol{R}) = T_{n_{\mathrm{i}},v}\left(\begin{bmatrix} g_v(z_{\mathrm{i}1}-\mu_{\mathrm{i}1}) \\ \vdots \\ g_v(z_{\mathrm{i}n_{\mathrm{i}}}-\mu_{\mathrm{i}n_{\mathrm{i}}}) \end{bmatrix} \middle| \boldsymbol{0},\boldsymbol{R}\right) \times \prod_{t=1}^{n_{\mathrm{i}}} \frac{\mathscr{L}(z_{\mathrm{i}t}|\mu_{\mathrm{i}t})}{T_{1,v}(g_v(z_{\mathrm{i}t}-\mu_{\mathrm{i}t})|0,1)} \tag{1.1}$$

where

$$T_{n_{\mathrm{i}},v}(\boldsymbol{t}_{\mathrm{i}}|\boldsymbol{\mu}_{\mathrm{i}},\boldsymbol{\Sigma}) = \left(\frac{\Gamma(\frac{v+n_{\mathrm{i}}}{2})}{\Gamma(\frac{v}{2})(v\boldsymbol{\pi})^{\frac{q}{2}}|\Sigma|^{\frac{1}{2}}}\right) \times \left(1+\frac{1}{v}(\boldsymbol{t}_{\mathrm{i}}-\boldsymbol{\mu}_{\mathrm{i}})^T\boldsymbol{\Sigma^{-1}}(\boldsymbol{t}_{\mathrm{i}}-\boldsymbol{\mu}_{\mathrm{i}})\right)^{-\frac{v+n_{\mathrm{i}}}{2}}$$

$$g_v(x) = F_v^{-1}\Big(\frac{\mathrm{e}^x}{1+\mathrm{e}^x}\Big)$$

$$F_v^{-1}: \text{ Inverse CDF of standard univariate T}$$

for some location parameter $\boldsymbol{\mu}_{\mathrm{i}}$.

The motivation for choosing the T distribution stems from work done by Albert and Chib [25] in that by using the degrees of freedom $v = 7.3$ and $\tilde{\sigma}^2 = \frac{\pi^2(v-2)}{3v}$, the standard T distribution greatly approximates the logistic distribution, i.e. $T_{1,v}(.|x_{\mathrm{i}}\beta,\tilde{\sigma}^2) \underset{\text{approx}}{\sim} \mathscr{L}(.|x_{\mathrm{i}}\beta,1)$. In following this approximation, exact inferences on $\boldsymbol{\pi(\beta,R)}$ can then be obtained via Importance Sampling [26].

The objective is to sample from the posterior $\boldsymbol{\pi(\beta,R,\phi,z|y)}$. We assume the following priors and latent configuration from Albert and Chib [25] and O'Brien and Dunson [4]

$$\boldsymbol{\beta} \sim N_p\Big(\boldsymbol{\beta}|\boldsymbol{\beta_0},\boldsymbol{\Sigma}_\beta\Big)$$

$$\boldsymbol{R} \sim \boldsymbol{\pi}(\boldsymbol{R})$$

$$\phi_{\mathrm{i}}|\boldsymbol{\beta},\boldsymbol{R} \sim \Gamma\Big(\phi_{\mathrm{i}}|\frac{v}{2},\frac{v}{2}\Big), \text{ where } v = 7.3$$

$$\boldsymbol{\pi}(\boldsymbol{z}_{\mathrm{i}}|\boldsymbol{\beta},\boldsymbol{R},\boldsymbol{\phi}) \sim N_{n_{\mathrm{i}}}\left(\boldsymbol{z}_{\mathrm{i}}\middle|\boldsymbol{x}_{\mathrm{i}}\boldsymbol{\beta},\frac{\tilde{\sigma}^2}{\phi_{\mathrm{i}}}\boldsymbol{R}_{\mathrm{i}}\right), \text{ where } \tilde{\sigma}^2 = \frac{\boldsymbol{\pi}^2(v-2)}{3v}$$

$$\boldsymbol{\pi(y_{\mathrm{i}}|\beta,R,\phi,z)}: \text{ truncation of } \boldsymbol{z}_{\mathrm{i}} \text{ based on the value of observed } \boldsymbol{y}_{\mathrm{i}}$$

Here $\boldsymbol{\pi}(\boldsymbol{R})$ is any distribution with support on the space of correlation matrices. Further, we assume a diffuse normal prior for $\boldsymbol{\beta}$ [4]. This translates to $\boldsymbol{\Sigma}_{\beta}^{-1}$ being a matrix with all 0 elements, i.e. a precision matrix full of zeroes.

To sample from the joint posterior $\boldsymbol{\pi}(\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{\phi}, \boldsymbol{z}|\boldsymbol{y})$, we use the Gibbs sampler from O'Brien and Dunson [4] such that at iteration (t)

1. For $\mathrm{i} = 1, \ldots, N$, sample from the posterior of $\boldsymbol{z}_{\mathrm{i}}$

$$\underset{n_{\mathrm{i}} \times 1}{\boldsymbol{z}_{\mathrm{i}}}^{(t)} | \boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\beta}, \boldsymbol{R}_{\mathrm{i}}, \phi_{\mathrm{i}}, \boldsymbol{x}_{\mathrm{i}} \sim N_q \left(\boldsymbol{x}_{\mathrm{i}} \ \boldsymbol{\beta}^{(t-1)}, \frac{\tilde{\sigma}^2}{\phi_{\mathrm{i}}^{(t-1)}} \boldsymbol{R}_{\mathrm{i}}^{(t-1)} \right)$$

   where z_{ij} is truncated above zero if $y_{\mathrm{ij}} = 1$ and below zero if $y_{\mathrm{ij}} = 0$.

2. For $\mathrm{i} = 1, \ldots N$, sample from the posterior of scalar $\phi_{\mathrm{i}}^{(t)}$

$$\begin{aligned} &\phi_{\mathrm{i}}^{(t)} | \boldsymbol{z}_{\mathrm{i}}, \boldsymbol{y}_{\mathrm{i}}, \boldsymbol{\beta}, \boldsymbol{R}_{\mathrm{i}}, \boldsymbol{x}_{\mathrm{i}} \\ &\quad \sim \Gamma \left(\frac{1}{2}(v + n_{\mathrm{i}}), \frac{1}{2} \left(v + \frac{1}{\tilde{\sigma}^2} \right) \left(\boldsymbol{z}_{\mathrm{i}}^{(t)} - \boldsymbol{x}_{\mathrm{i}} \boldsymbol{\beta}^{(t-1)} \right)^T (\boldsymbol{R}_{\mathrm{i}}^{(t-1)})^{-1} \left(\left(\boldsymbol{z}_{\mathrm{i}}^{(t)} - \boldsymbol{x}_{\mathrm{i}} \boldsymbol{\beta}^{(t-1)} \right) \right) \right) \end{aligned}$$

3. Sample from the posterior of $\boldsymbol{\beta}^{(t)}$

$$\begin{aligned} \boldsymbol{\beta}^{(t)} | \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{R}, \boldsymbol{\phi}, \boldsymbol{x} &\sim N_p(\tilde{\boldsymbol{\mu}}_{\beta}, \tilde{\boldsymbol{\Sigma}}_{\beta}) \\ \tilde{\boldsymbol{\Sigma}}_{\beta} &= \left(\Sigma_{\beta}^{-1} + \frac{1}{\tilde{\sigma}^2} \sum_{\mathrm{i}=1}^{n} \phi_{\mathrm{i}}^{(t)} \boldsymbol{x}_{\mathrm{i}} (\boldsymbol{R}^{(t-1)})^{-1} \boldsymbol{x}_{\mathrm{i}} \right)^{-1} \\ \tilde{\boldsymbol{\mu}}_{\beta} &= \tilde{\Sigma}_{\beta} \left(\Sigma_{\beta}^{-1} \beta_0 + \frac{1}{\tilde{\sigma}^2} \sum_{\mathrm{i}=1}^{n} \phi_{\mathrm{i}}^{(t)} \boldsymbol{x}_{\mathrm{i}} (\boldsymbol{R}^{(t-1)})^{-1} \boldsymbol{z}_{\mathrm{i}}^{(t)} \right) \end{aligned}$$

4. Sample from the posterior of $\boldsymbol{R}^{(t)}$

$$\boldsymbol{\pi}(\boldsymbol{R} | \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{x})$$

The BMLR procedure is summarized in Figure 1.2.

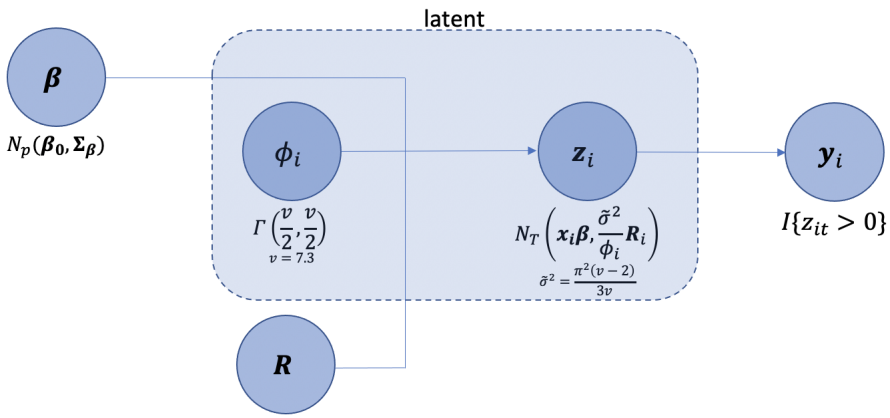


Figure 1.2. Summary of BMLR procedure

To sample from $\boldsymbol{\pi}(\boldsymbol{R}|\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{x})$, we use the following protocol from O'Brien and Dunson [4]

1. Sample each $n_{\mathrm{i}}^* = \frac{n_{\mathrm{i}}(n_{\mathrm{i}}-1)}{2}$ from

$$\text{unique } \tilde{\boldsymbol{R}} \sim N_{n_{\mathrm{i}}^*}(\text{unique } \boldsymbol{R}^{(t-1)}, \boldsymbol{\Omega})$$

   where $\boldsymbol{\Omega}$ is a tuning parameter chosen via experimentation.

2. If this new $\tilde{\boldsymbol{R}}$ is positive definite, then set $\boldsymbol{R}^{(t)} = \tilde{\boldsymbol{R}}$ with probability

$$\min\left\{1, \frac{\boldsymbol{\pi}(\tilde{\boldsymbol{R}}) \prod_{\mathrm{i}=1}^{N} N_{n_{\mathrm{i}}}\left(\boldsymbol{z}_{\mathrm{i}}^{(t)} \middle| \boldsymbol{x}_{\mathrm{i}}\boldsymbol{\beta}^{(t)}, \frac{\tilde{\sigma}^2}{\phi_{\mathrm{i}}^{(t)}}\tilde{\boldsymbol{R}}\right)}{\boldsymbol{\pi}(\boldsymbol{R}^{(t-1)}) \prod_{\mathrm{i}=1}^{N} N_{n_{\mathrm{i}}}\left(\boldsymbol{z}_{\mathrm{i}}^{(t)} \middle| \boldsymbol{x}_{\mathrm{i}}\boldsymbol{\beta}^{(t)}, \frac{\tilde{\sigma}^2}{\phi_{\mathrm{i}}^{(t)}}\boldsymbol{R}^{(t-1)}\right)}\right\}$$

   Otherwise, set $\boldsymbol{R}^{(t)} = \boldsymbol{R}^{(t-1)}$

3. If this new $\tilde{\boldsymbol{R}}$ is not positive definite, then set $\boldsymbol{R}^{(t)} = \boldsymbol{R}^{(t-1)}$.

As shown in Figure 1.1, with larger number of repeated measures, this protocol becomes far too restrictive, resulting in a sampler that does not explore the correlation space effectively. Before we propose our prior to circumvent this, we first look at Partial Autocorrelations.

1.3 Methodology

1.3.1 Partial Autocorrelation

Partial autocorrelations of an AR(k) process is 0 at lag $k+1$ and greater [27]. Thus we only need the first k partial autocorrelations to describe an AR(k) process [28]. We then use the recursive algorithm to calculate the marginal autocorrelations from the partial autocorrelations [29]. Following the notation in Joe [30], for $\mathrm{j} = 1, \ldots, n_\mathrm{i} - k$, $k = 1, \ldots, K-1, K \leq \min_\mathrm{i}\{n_\mathrm{i}\}$, once we have the partial autocorrelation $\rho_{\mathrm{j},\mathrm{j}+k}$, we can recursively populate the marginal correlation matrix as follows:

$$r_{\mathrm{j},\mathrm{j}+k} = r_1'(\mathrm{j},k) R_2(\mathrm{j}.k)^{-1} r_3(\mathrm{j},k) + D_{\mathrm{j}k}\rho_{\mathrm{j},\mathrm{j}+k} \tag{1.2}$$

where

$$\begin{aligned}
r_1'(\mathrm{j},k) &= (r_{\mathrm{j},\mathrm{j}+k}, \ldots, r_{\mathrm{j},\mathrm{j}+k-1}) \\
r_3'(\mathrm{j},k) &= (r_{\mathrm{j}+k,\mathrm{j}+1}, \ldots, r_{\mathrm{j}+k,\mathrm{j}+k-1}) \\
R_2(\mathrm{j},k) &= \boldsymbol{R}_{[\mathrm{j}+1,\ldots,\mathrm{j}+k-1:\mathrm{j}+1,\ldots,\mathrm{j}+k-1]} \\
D_{\mathrm{j}k} &= [1 - r_1'(\mathrm{j},k) R_2(\mathrm{j},k)^{-1} r_1(\mathrm{j},k)]^{\frac{1}{2}} [1 - r_3'(\mathrm{j},k) R_2(\mathrm{j},k)^{-1} r_3(\mathrm{j},k)]^{\frac{1}{2}}
\end{aligned}$$

Thus, we simply need to model the partial autocorrelations $\boldsymbol{\rho}$, and we can recover the correlation matrix $\boldsymbol{R}$.

The partial autocorrelations are constrained on $(-1, 1)$, making posterior sampling somewhat confining. To circumvent this, we do a Fisher transformation (or atanh()) on the partial autocorrelations [31]. Now if we assume for person i with repeat measures n_i that $\boldsymbol{\rho}$ is the vector of partial autocorrelations of length $k = 1, \ldots, K-1, K \leq \min_\mathrm{i}\{n_\mathrm{i}\}$. Then $\rho_k^z = \mathrm{atanh}(\rho_k) \underset{iid}{\sim} \mathrm{logistic}\Big(\mu = 0, s = \frac{1}{2}\Big)$.

1.3.2 BGLAM

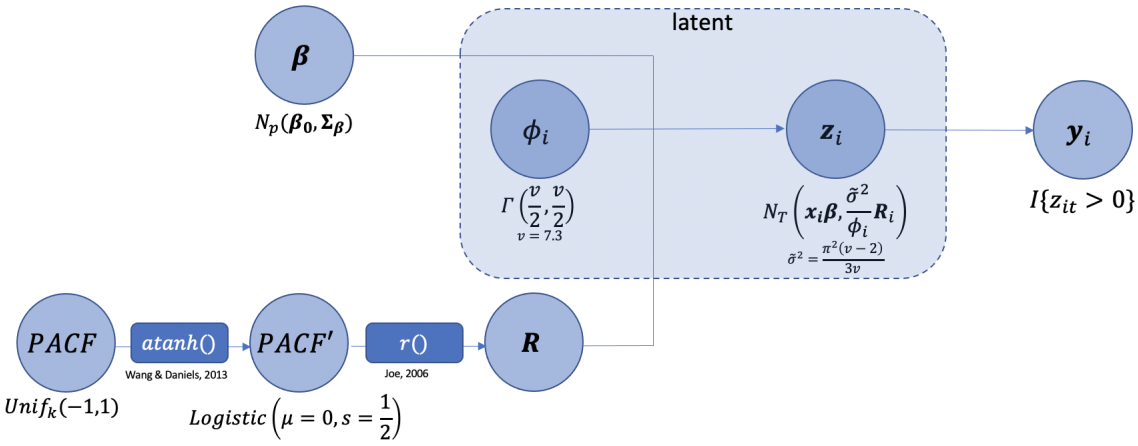


Figure 1.3. Summary of the BGLAM procedure

Our proposed model BGLAM is different from the original BMLR in that we have a prior for $\boldsymbol{R}$ that can better accommodate an autoregressive structure. Sampling from the posterior $\pi(\boldsymbol{R}|\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{x})$ is equivalent to sampling from the posterior of $\pi(\boldsymbol{\rho}|\boldsymbol{z}, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{x})$. Define the function $r(.)$ to be the recursive process to go from the partial to marginal autocorrelations [32]. Then the posterior of $\boldsymbol{\rho}$ decomposes to

$$\pi(\boldsymbol{\rho}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{x}) \propto \left[\prod_{\mathrm{i}=1}^{N} N_{n_\mathrm{i}}\left(\boldsymbol{z}_\mathrm{i}|\boldsymbol{x}_\mathrm{i}\boldsymbol{\beta}, \frac{\tilde{\sigma}^2}{\phi_\mathrm{i}} r(\tanh(\boldsymbol{\rho}^z))\right)\right]\left[\prod_{k=1}^{K} Logistic\left(\rho_k^z|0, \frac{1}{2}\right)\right] \tag{1.3}$$

We will use the Metropolis Hastings algorithm with a Normal proposal distribution to sample from $\pi(\boldsymbol{\rho}|\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{x})$. By experimentation, we set the tuning parameter for the proposal distribution as $\sigma_p^2 = 0.001$.

1.3.3 Weighted Posterior Predictive P-Values

We construct posterior predictive p-values (PPP) to draw inference from our Bayesian model on the parameters of interest. Following Gelman [33, p. 145], let TS(.) be some test statistic function, $y^{(rep)}$ be the replicated data and $\theta = (\beta, \rho, z)$. Specifically, we define TS(.)

as the difference in proportion of success of 1's of each subject between the treatment and control group. By definition, $PPP = \int I\{\text{TS}(y^{(rep)}, \theta) > \text{TS}(y, \theta)\}\boldsymbol{\pi}(y^{(rep)}, \theta|y)dy^{(rep)}d\theta$.

Since we are using Importance Sampling to get exact inferences, we need to adjust our PPP using the importance weights. By defining $\boldsymbol{\pi}^*(\theta|y), \boldsymbol{\pi}(\theta|y)$ and $\boldsymbol{\pi}(y^{(rep)}|\theta)$ as the approximate likelihood, true likelihood and simulated outcomes based on the true model, the weighted PPP is now simply

$$PPP_{impt} = \mathbb{E}_{y^{(rep)},\theta|y}\left[I\{\text{TS}(y^{(rep)}) > \text{TS}(y)\}\left[\frac{\boldsymbol{\pi}(\theta|y)}{\boldsymbol{\pi}^*(\theta|y)}\right]\right] \tag{1.4}$$

As $\boldsymbol{\pi}(\theta|y)$ and $\boldsymbol{\pi}^*(\theta|y)$ are close approximates of each other, their ratio is quite stable for moderate to large number of repeated measurements.

1.3.4 Marginal Likelihood

A typical goodness of fit measure used is the marginal likelihood [21], [34], since it captures the probability of observing the data across the input parameter space. However, there is no closed form for calculating the log likelihood of our BGLAM model. Specifically, we want to calculate the following:

$$\Pr(\boldsymbol{Y}_{\text{i}} = \boldsymbol{y}_{\text{i}}|\boldsymbol{x}_{\text{i}}, \boldsymbol{\beta}, \boldsymbol{R}) = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\left\{\prod_{\text{j}=1}^{n_{\text{i}}} I\{z_{\text{ij}} > 0\}^{y_{\text{ij}}} I\{z_{\text{ij}} \leq 0\}^{1-y_{\text{ij}}}\right\}\mathscr{L}_{n_{\text{i}},v}(\boldsymbol{z}_{\text{i}}|\boldsymbol{x}_{\text{i}}\boldsymbol{\beta}, \boldsymbol{R})d\boldsymbol{z}_{\text{i}} \tag{1.5}$$

To resolve this, we reparameterize the integrals above as follows:

$$\Pr(\boldsymbol{Y}_{\text{i}} = \boldsymbol{y}_{\text{i}}|\boldsymbol{x}_{\text{i}}, \boldsymbol{\beta}, \boldsymbol{R}) = \int_{\Omega_{\text{i}1}}\dots\int_{\Omega_{\text{i}n_{\text{i}}}}\mathscr{L}_{n_{\text{i}},v}(\boldsymbol{z}_{\text{i}}|\boldsymbol{x}_{\text{i}}\boldsymbol{\beta}, \boldsymbol{R})d\boldsymbol{z}_{\text{i}} \tag{1.6}$$

where

$$\Omega_{\text{ij}} = \begin{cases}(-\infty, 0), \text{ if } y_{\text{ij}} = 0\\ (0, \infty), \text{ if } y_{\text{ij}} = 1\end{cases}$$

With the above likelihood, we can approximate the marginal likelihood by averaging across the input parameter space.

1.3.5 Deviance Information Criterion

The DIC [35]–[37] is a widely applicable Bayesian measure used for model comparison. Deviance is defined as $D(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{X}) = -2\log\left(Pr(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{X})\right)$. The posterior mean deviance is $\bar{D}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{X}) = \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{R}|\boldsymbol{Y},\boldsymbol{X}} D(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{X})$. The effective number of parameters is then calculated via $p_D = \bar{D}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{X}) - D(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{X})$. Finally, the DIC is calculated by the expression $DIC = \bar{D}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{R}, \boldsymbol{X}) - p_D$. DIC is similar to the Akaike Information Criterion (AIC) in that AIC is based on the maximum likelihood estimate of the parameters while DIC is based on the posterior summary. Compared to other Bayesian measures of fit, the DIC can be computed with relative ease and is numerically stable [38].

1.4 Simulation Studies

1.4.1 AR(1) Simulation

A total of 1000 data replicates were generated for 100 subjects, each with 30 measurements. Each simulated subject would have a continuous covariate sampled from a standardized normal distribution. We have three estimable parameters: $(\beta_0, \beta_{\text{trt}}, \beta_1)$ where β_0 is the intercept and β_1 is the coefficient associated with the generated continuous covariate, and the partial autocorrelations ρ. Based on an Intention-To-Treat(ITT) analysis on our preliminary data, we set $(\beta_0, \beta_1) = (-2, 1.5)$. The β_{trt} parameter combinations are chosen from $(0, 0.25, 0.5)$ with the AR(1) parameter from $(-0.7, -0.5, -0.3, 0.3, 0.5, 0.7)$. We assess model performances based on Type-1 error rates, power and coverage probabilities. We simulate our data from three different processes: T-copulas, Normal-copulas and GLMM. Normal-copulas were chosen based on [19]. Data were also generated via GLMM since we are comparing our BGLAM procedure to GLMM and GEE. We compare our BGLAM model to the GLIMMIX procedure on both the G-side and R-side as well as GEE via the gee package [39] in R. Simulation settings for AR(1) are summarized in Table 1.1. The proposed model

BGLAM is depicted by the black solid line in all AR(1) simulation results. SAS scripts the GLIMMIX-G and GLIMMIX-R are provided in Appendix A.A.

Table 1.1. Input parameters for AR(1) simulations with 150 subjects and 30 repeated measurements each.

beta_trt	AR(1)	Data Generator	method
0 0.25 0.5	0.7 0.5 0.3 -0.3 -0.5 -0.7	GLMM N-Copula T-Copula	BGLAM GLIMMIX-G GLIMMIX-R GEE

The power of the BGLAM is better, if not just as good as GEE, GLIMMIX (both G and R sides) using Kenward Rogers adjustment. This behavior is more apparent for positive AR(1) values or when there is more separation in the data (Figure A.1). Although it may seem that GLIMMIX may have higher power at times, this comes at the cost of elevated Type I error rates (around 0.9 at $\alpha = 0.05$). This is especially apparent in Figure *A*.12 when AR(1)=0.7. In addition, the figure also shows how BGLAM outperforms GEE and GLIMMIX in terms of power, even when the data were generated via GLMM. Further, BGLAM captures the correct AR(1) estimates under the T-copula process for all AR(1) settings (Figures A.7-A.12). The AR(1) estimates from GLIMMIX-G are consistently biased, even when the data were generated via GLMM. A limitation of GEE models in repeated measures designs is that they do not have standard error estimates for the AR(1) parameters. Bootstrapping is not feasible as well since this would result in duplicate subjects across time, yielding estimation errors due to perfect autocorrelations. Thus, GEE coverage for the partial autocorrelation estimates will be absent.

When data were generated via GLMM, there is a mismatch between the generated data and our proposed BGLAM, resulting in undercoverage A.1-A.6, leftmost column. The GLIMMIX procedure still suffers from treatment undercoverage even when the data were generated

via GLMM (Figures A.7-A.12, leftmost column). This is because the AR(1) estimates tend to swing from extreme positive to extreme negative AR(1) estimates. Specifically, the treatment coverage when AR(1)=0.7 (Figure A.1) and AR(1)=-0.7(Figure A.6) is not as bad as the intermediate AR(1) settings in (Figures A.2-A.5). Simulation results for the remaining AR(1) settings are available in the Supplementary Materials in Section A.

1.4.2 AR(2) Simulation

Table 1.2. Setup for AR(2) simulations.

$beta_{\text{trt}}$	AR(2)	Data Generator	Method
0 0.25 0.5	(-0.7, -0.5) (-0.7, 0.5) (0.7, -0.5) (0.7, 0.5) (-0.5, -0.3) (-0.5, 0.3) (0.5, -0.3) (0.5, 0.3)	T-Copula	BGLAM-AR(1) BGLAM-AR(2) BGLAM-AR(3) GEE-AR(2)

As before, simulations are set for 100 subjects, each with 30 measurements. However we now have three estimable parameters: $(\beta_0, \beta_{\text{trt}}, \beta_1)$ and a vector of PARs whose length depends on the k in the AR(k). The β parameter settings are similar as before, just that now AR(2) $= \{(-0.7, -0.5), (-0.7, 0.5), (0.7, -0.5), (0.7, 0.5), (-0.5, -0.3), (-0.5, 0.3), (0.5, -0.3), (0.5, 0.3)\}$. Simulation settings for AR(2) are summarized in Table 1.2. We compared the correctly specified BGLAM-AR(2) model against the incorrectly specified BGLAM-AR(1), BGLAM-AR(3) and GEE-AR(2), or GEE with an AR(2) specification, using [39]. We assess model performances based on power, bias, coverage probabilities and standard errors of the estimates over the 1000 data replicates. Further, we assess model fit using the Deviance Information Criterion (DIC) and the Marginal Likelihood. The model fits are supplemented with the 95% credible intervals for the PAR(k) estimates.

Simulation results show that a correctly specified BGLAM-AR(k) has better, if not comparable power. The BGLAM with the correctly specified AR(2) structure has better power (Figure A.13) in most settings. Even with the correct AR(2) structure, BGLAM outpowers GEE (Figure A.13). This is not always true when the autocorrelation is positive and sticky or does not decay as quickly. For example, when the true AR(2)=(0.5, 0.3), the power for BGLAM-AR(2) is slightly lower than that of BGLAM-AR(1). This is because the binary outcomes are not only highly positively correlated, the strength of the correlation lingers longer and decays more slowly as compared to an AR(1)=0.5 setting, yielding binary outcomes that are mostly 1's or 0's. The power disparity is even more prominent when AR(2)=(0.7, 0.5). At this positive correlation strength however, all different AR(k) settings for BGLAM and GEE with an AR(2) setting suffer lower power.

Further, a correctly specified BGLAM-AR(k) has better coverage with lower bias and standard error. Although the correct BGLAM-AR(2) has low bias, all BGLAM settings suffer from lower than favorable bias when the true AR(2)=(-0.7, 0.5) and (0.7, 0.5). This goes in tandem with the coverage of the corresponding posterior PARs in Figure A.14, where the PAR posterior means are consistently higher than the true PARs. Similarly, treatment coverage hovers around the 0.95 threshold. We do see that coverage for BGLAM-AR(1) is sometimes lower and even dips to as low as 0.7. This is not true for BGLAM-AR(3), since the last PAR is often close to 0, yielding an induced marginal correlation that follows the correct AR(2) (Figure A.14). GEE-AR(2) does have the lowest bias and consistent coverage of 0.95 throughout all settings. This is expected since GEEs are robust and not sensitive to model misspecification. Finally, the standard errors are fairly similar across the different simulation settings, with GEE being consistently higher than the rest.

The correct model is selected using a combination of DIC, marginal log likelihood and the CI of the last PAR of the corresponding model. The differences between the fit measures for a given simulation setting are quite small, often to one decimal place. This is because differences induced by the varying AR(k) structures are fairly subtle. For example, the DIC and marginal log likelihood select the correct AR(2) model when the true AR(2)=(-0.7, 0.5) in Table 1.4. However, when AR(2)=(0.7, -0.5), the AR(3) model has the lowest DIC and Marginal LL. Note that the differences of these fit measures with that of AR(2) is to one or

second decimal places. The zero coverage probability for the last PAR of the AR(3) models are around 0.7. Favoring parsimony, we would opt for the AR(2) model. The converse is also true. For instance, although the AR(1) models have the lowest DIC and Marginal LL in some settings, the differences between these fit measures with that of AR(2) is to one or second decimal places. Similarly, the zero coverage probability for the last PAR of the AR(1) and AR(2) models are 0 while that of AR(3) is around 0.9, indicating that an AR(2) would be the ideal model given how close these fit measures are. Note that BGLAM-AR(1) has markedly better DIC and marginal LL measures than BGLAM-AR(2) when AR(2)=(0.7, 0.5). As before, this correlation setting would yield data that are mostly either 1's pr 0's, compelling our model to favor parsimony in parameterizing the AR(k) structure with limited information to work on.

PAR	ar	Marginal Log Likelihood	DIC	Last PAR Zero-CP
(-0.7,0.5)	1	-33.191 (3.685)	67.314 (7.508)	0
	2	-30.553 (2.851)	61.513 (5.733)	0
	3	-30.669 (2.88)	61.798 (5.822)	0.788
(-0.5,0.3)	1	-29.13 (2.491)	58.463 (4.983)	0
	2	-29.145 (2.47)	58.528 (4.943)	0.003
	3	-29.173 (2.496)	58.583 (4.986)	0.9
(-0.7,-0.5)	1	-29.176 (2.477)	60.683 (4.818)	0
	2	-29.618 (2.576)	61.48 (5.018)	0
	3	-29.329 (2.502)	61.014 (4.848)	0.519
(-0.5,-0.3)	1	-29.153 (2.46)	60.613 (4.784)	0
	2	-29.162 (2.472)	60.505 (4.808)	0.001
	3	-29.161 (2.467)	60.53 (4.787)	0.892
(0.7,-0.5)	1	-30.023 (2.958)	60.353 (5.912)	0
	2	-29.9 (2.974)	60.069 (5.94)	0
	3	-29.855 (2.99)	59.969 (5.967)	0.696
(0.5,-0.3)	1	-29.475 (2.633)	59.233 (5.267)	0
	2	-29.366 (2.637)	58.993 (5.268)	0
	3	-29.34 (2.64)	58.937 (5.267)	0.874
(0.7,0.5)	1	-29.851 (3.33)	66.528 (6.863)	0
	2	-31.348 (3.353)	73.573 (7.294)	0
	3	-31.539 (3.347)	74.538 (7.395)	0.556
(0.5,0.3)	1	-29.103 (2.751)	62.152 (5.417)	0
	2	-29.342 (2.748)	63.447 (5.445)	0
	3	-29.354 (2.751)	63.47 (5.472)	0.908

Figure 1.4. Fit measures and standard errors for the AR(2) simulation study for 1000 data replicates generated via T-copula. Highlighted grey rows are the correct models. The difference between the fit measures between the all AR(k) settings are to one decimal place. The exception to this is when the true PAR = (-0.7, 0.5), where the difference between the fit measures between AR(1) and AR(2) are much prominent. Model selection is further supplemented by the high zero coverage probability for the last PAR of AR(3).

1.5 Application: PTSD Clinical Trial

1.5.1 Description Of Study

We demonstrate the application of our BGLAM methodology by considering an NIH-funded clinical trial (#NCT03245814) conducted by Dr. Marguerite O'Haire on the effectiveness of service dogs for United States military Veterans diagnosed with post traumatic stress disorder (PTSD). This study consists of two treatment groups: Veterans assigned a service dog (the active treatment of interest), and Veterans who were on a waitlist to receive a service dog (control group). The organization that provided service dogs to Veterans in this study is K9's For Warriors.

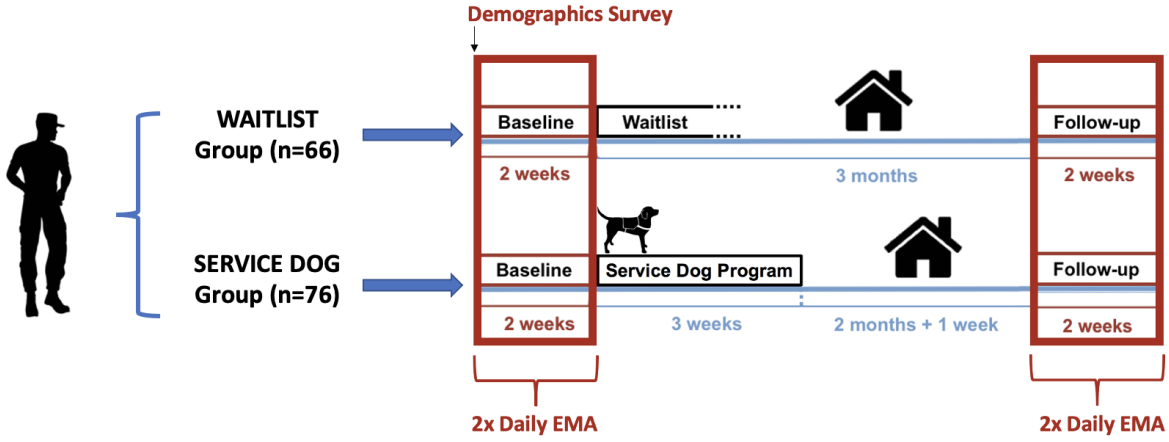


Figure 1.5. The design protocol for Veterans in the Service Dog (bottom with 76 Veterans) and waitlist group (top with 66 Veterans). Veterans fill out the demographics survey at the beginning of their study period. They then receive pings to complete short EMA questionnaires on their mobile devices twice daily at baseline and follow-up for two weeks. Specifically, Veterans are asked to indicate where they are at that present moment. Between baseline and follow-up, Veterans assigned to the Service Dog group undergo a training course with their assigned service dogs, followed by a live-in period with the dogs in their respective homes for a total of 3 months. Veterans assigned to the waitlist group will undergo the same procedure after the end of their respective study periods.

Three phases are involved in this study: (1) a period of two weeks in which baseline data on the participants are collected, (2) a subsequent period of three months in which only the

treatment group engages in a training course with their assigned service dogs followed by a live-in period with the dogs in their own homes (Figure 1.5), and (3) a final period of two weeks in which follow-up data are collected. Data were collected following Ecological Momentary Assessment (EMA) protocols, which involve repeated sampling of participants' current experiences and behaviours in real time in their innate environment [40]. Further, timing of the sampling was randomized to prevent anticipatory responses and capture a breadth of daily experiences. In an EMA procedure, participants receive notifications to complete short questionnaires on their mobile phones twice daily during the baseline and follow-up periods. Each questionnaire asks participants to indicate where they are at the present moment. In the analysis, the binary repeated measures of interest consist of the indicators for whether or not a participant is at home during particular points in time during the follow-up period. The data consists of 142 Veterans, each with approximately 28 repeated measurements during follow-up.

The objective in analyzing this repeated measures study is to determine the effect of a service dog on the probability that a Veteran will be out of their home, accounting for baseline covariates and time dependencies. Previous literature suggests that pairing of a Veteran with a service dog for PTSD enables the Veteran to be more involved in their communities and go out in public more often [41], [42]. However, the majority of evidence to support this is qualitative in nature. Thus, the present study adds to the current literature by quantitatively analyzing the odds of a Veteran being out of their home with the addition of a service dog in comparison to a waitlist control group. Results of this analysis may assist service dog providers and mental health practitioners by providing additional insight into the service dog intervention.

A reasonable autocorrelation structure in this context is the autoregressive (AR) process, which has the correlation in the outcomes between any two points in time diminishing as a function of their separation in time. To specify the order of the AR process we note that as data are collected twice daily, a subject may be less likely to leave home with their service dog if they have done so at the time point immediately before the current time point, and that they are more likely to do so during the time point immediately after the current time point.

1.5.2 Model

The outcome is a binary 0-1 variable: the outcome is 1 if the participant indicated that they were at home at the time of the questionnaire notification and 0 otherwise. Specifically, the outcome is 0 if the participant indicated that they were at work or school, in transit, at a doctor or therapy, in an indoor public space, in an outdoor public space, at a friend or family's house, or if they selected "Other" in their questionnaire response. We are modelling the odds of the participants being home. Three predictors are used in our model: treatment assignment (service dog vs waitlist group), age and proportion of time Veterans were home at baseline. Here age was included as a demographic variable that may be related to leaving their homes.

There are four separate components that go into the correlation structure for our model: $\boldsymbol{\rho}_{\text{weekday}}$; a vector of partial autocorrelations that defines an AR(k) structure for weekday, defined as Monday through Thursday, $\boldsymbol{\rho}_{\text{weekend}}$; a vector of partial autocorrelations that defines an AR(k) structure for weekend, defined as Friday through Sunday, $\rho_{\text{Thur-Fri}}$; a correlation term that captures correlation between Thursday evening and Friday morning, and $\rho_{\text{Sun-Mon}}$; a correlation term that captures correlation between Sunday evening and Monday morning. The motivation behind this correlation structure stems from the weekend effect, which states that human behavior adheres to a cyclic rhythm that fluctuates between weekdays and weekends [43]. Weekend effects are consistent with other daily diary studies of mood [44]–[46].

Thus our model should be able to accommodate for this intricacy. We follow the definition of weekends from Ryan, Bernstein, and Brown [47] which clusters Friday through Sunday as weekends and Monday through Thursday as Weekdays. However, these weekday and weekend unique AR(k) structures are not stand-alone units and should be able to transition from one to the other interchangeably. Therein, we introduce the two scalar correlation inputs $\rho_{\text{Thur-Fri}}$ and $\rho_{\text{Sun-Mon}}$, where the former captures the correlation between weekday and weekend and latter captures the correlation between each week. We ran our specified model on 4 different AR(k) settings for 10250 iterations and a burn-in of 1250: AR(1) for

both weekday and weekend, AR(2) for both weekday and weekend, AR(3) for both weekday and weekend, and a custom AR(2) for weekday and AR(1) for weekend.

1.5.3 Results

Table A.1 summarizes the DIC and marginal loglikelihood. We see that the custom AR(k) setting has the best measure on all fit measures, albeit marginal. Table 1.3 provides the posterior summaries for the treatment and covariates. Posterior means for the different PARs for all the AR(k) settings are summarized in Table A.1.

Further, we ran a sensitivity analysis to see how the treatment effect changes over the varying AR(k) settings. Table A.1 summarises this information and shows that the inference on the treatment effect remains insignificant. Table A.1 lists the posterior summaries for the four different partial autocorrelation components for the selected model: the weekday correlations are governed by an AR(2)=(0.21, 0.40) process while the weekend correlation is modelled via an AR(1)=0.14 process. The correlation estimates for weekday and weekend partitions are both 0.1.

We conclude that the treatment assignment (service dog vs waitlist group) has no significant impact on the odds of a participant being inside their homes. However, it is noteworthy that the 95% credible interval does not cover the zero value for the baseline mean covariate. This implies that participant behavior at baseline may have an association with their behavior at followup, i.e. the proportion of time participants spend being inside their homes at baseline is associated with the odds being of being home at followup.

Table 1.3. Posterior Summaries of parameters for the custom 2 weekday and 1 weekend AR(k) setting.

Variable	Posterior Mean	SD	95% Lower CI	95% Upper CI
Intercept	0.77	0.09	0.58	0.95
Treatment	0.08	0.14	-0.2	0.36
Age	0.10	0.05	-0.01	0.2
Baseline Mean	-0.26	0.05	-0.37	-0.16

1.6 Discussion

In applying the BGLAM model to the NIH service dog study, we conclude that there is no significant association between being assigned a service dog and being outside of the house at 3-months followup. Instead, there is a larger relationship between baseline and follow-up at the individual level, such that participants who were more likely to be at home at baseline were still more less to be at home at 3-months follow-up. Thus, there may be robust individual differences in daily structures, routines, and propensity to leave the house that are stronger than the measured treatment effect in this study.

However, these findings do not paint a holistic picture of how service dogs can affect Veterans with PTSD. Despite prior literature suggesting that service dogs may aid Veterans in leaving their homes [48]–[51], this may not be represented consistently enough in the intervention for patterns to be detected. For example, this qualitatively reported effect could be tied to significant confounders not included in the present quantitative analyses. Additionally, the outcome survey is asking whether the Veteran was at home at only the specific time of notification. Thus, it could be that the participants were out of the house most of the time with the service dogs, but were home during the times they responded to the survey question. Further, being outside of the home is but one of the many measures clinicians use to assess the efficacy of a service dog in helping Veterans with PTSD.

Although an unstructured covariance assumption would be the most general one, this study is limited in the number of subjects relative to the number of repeat measures. For example, running GLIMMIX in SAS on a simulated AR(1) data with 28 repeat measures and 100 subjects took around 15 hours that resulted in no convergence. Running GENMOD procedure in SAS or the gee in R in would often return working correlation matrices that were not positive definite.

We would like to note that this is a marginal model that is unable to capture other sources of random variability, much like the GEE. A possible extension would be to augment the current GLMM model on STAN and supply a custom prior. Alternatively, one could explore the Polya-Gamma Random Variable [52] and include the priors for partial autocorrelations introduced in this paper.

One key feature of this BGLAM model is that it is a Bayesian model that can be used directly in causal models such as Principal Stratification [53] in longitudinal studies. Although these causal models have been implemented as such in [54] and [55], a flexible autoregressive model has never been used as of the submission of this paper. This would be especially useful in assessing causality of drug interventions with non-compliance in longitudinal studies [56].

Further, we note that the BGLAM model is not limited to an AR(k) process. For instance, a Toeplitz structure for the correlation matrix with varying bandwidths can be used in place of the AR(k) structure. Users simply need to swap out the prior configuration from an AR(k) process to sample from the posteriors of the Toeplitz entries of interest and populate the correlation matrix $\boldsymbol{R}$ from Figure 1.3.

Another possible extension is to set a prior on the order k in the AR(k) structure and infer it using a reversible-jump MCMC process [57]. This added feature would add significantly to the already intricate BGLAM sampler and warrants a separate work on its own. This possible extension is something we hope to investigate in the future.

In conclusion, the BGLAM is a Bayesian model that allows users to have varying nested AR(k) structures in the same model with effects and correlations interpretable on the log odds scale equipped with high power in longitudinal binary data.

2. GLAMRE: A GENERAL LOGISTIC AUTOREGRESSIVE MODEL WITH RANDOM EFFECTS FOR HETEROGENEOUS CORRELATED BINARY OUTCOMES

2.1 Introduction

Prior to modelling clustered binary data or those with hierarchical groupings, researchers will have to choose between a marginal or conditional model. Intuitively, marginal models can be seen as drawing inference on the macro or population level while conditional models do this on the micro level by incorporating local sources of variability. Marginal models include Generalized Estimating Equations [GEE; 5]–[8] and the Bayesian Multivariate Logistic Regression [4]. On the other hand, conditional models comprise Generalized Linear Mixed Models [GLMM; 9] and Bayesian Hierarchical Generalized Linear Models [33]. Having to choose one over the other, however, is still an open discussion [58]–[61].

A key advantage of marginal models is that parameter estimation in these models are less computationally demanding and more robust to model misspecifications relative to conditional models [5], [59], [62]. In the case of GEE, should the within-cluster correlation structure be misspecified, the marginal regression parameters will still be consistent, i.e. the sampling distribution of the estimator becomes increasingly emphasized around the true value [5], [63]. Unlike conditional models such as GLMMs, marginal models do not require distributional assumptions about random effects. As such, a common critique of GLMMs is their reliance on the normality of the random effects, rendering them more susceptible to violations of model assumptions. Although GLMMs are afforded more flexibility in modelling heterogeneous sources of variability, fitting these models run the risk of being computationally expensive. For instance, fitting a conditional model that incorporates spatial and/or temporal associations [12], [61] can be computationally consuming and would likely require a lot more data to be able to separately estimate these correlation structures. Further, users will have to choose from a gamut of methods to fit the GLMM such as penalized quasilikelihood [PQL; 64], adaptive Gauss-Hermite quadrature [GHQ; 13], Laplace Approximation

[23] or Bayesian approaches using Markov Chain Monte Carlo (MCMC) sampling procedures [65].

Unlike the marginal approach, conditional models can offer more insight into the layered sources of variability. In addition to being able to model complex layers of variability, a key advantage of conditional models concerns model selection. Specifically, the existence of a tractable true likelihood function. Although the Quasi-likelihood Information Criterion [QIC; 66] was proposed as a model selection tool in GEE models, applications show that QIC does not consistently select the correct model [67]. One could get MCMC approximates of other fit statistics such as the Deviance Information Criterion [DIC; 35] in Bayesian marginal models [4]. However, these can be computationally expensive. In addition, one can simulate the population model to get marginal inferences from a conditional model, assuming the distributions of the covariates are provided.

In this chapter, we will use the conditional approach to model the NIH Service Dog study from Section 1.5.1 for the following reasons: firstly, treatment assignment in this study was assigned by cohorts due to limited time and space. Therein lies a layer of random variability from cohort. It would be unwise to ignore this source of variability at the risk of having an inflated Type I error rates and biased parameter estimates. Secondly, since there are around 30 repeated measurements per subject in the collected data, using a marginal model that can only facilitate a correlation matrix [4] might be too restrictive. To that end, we will model the possible heterogeneity in the data over time by including a scale parameter for the time covariance matrix.

Despite the utilities of a conditional approach, existing frequentist-based statistical programs are not as computationally efficient and lack flexibility in modelling nested AR(k) processes. The GLIMMIX procedure in SAS®, which adopts a frequentist approach to modelling GLMM, takes around 15 hours to finish running a Binary Repeated Measures Model on a single simulated data set with 150 subjects and 2 predictors, each with 30 repeat measurements that follow an AR(1) structure alongside a random intercept for 25 levels for cohort using the Kenward-Roger adjustment. Its Bayesian analog, the BGLIMM procedure, can run this model and converge under in 5 minutes on the same machine. This is still unsatisfactory since both GLIMMIX and BGLIMM only allow for a single AR(1) process in any

single model, foregoing the capacity to uncover layers of layered autocorrelation, potentially biasing estimates and inflating Type I error rates due to misspecifications. This is especially relevant in studies with many repeat measurements where nested correlations are more likely to manifest by virtue of having more data to glean autocorrelations from.

Further, it is not straightforward to sample from conjugate posteriors for binary data, relative to normal data since the likelihood is not in a malleable form [68]. As such, procedures such as the Hamiltonian Monte Carlo [HMC; 69] or the Gammerman algorithm [70] are used to derive the relevant posteriors in non-normal data. Both HMC and the Gammerman algorithm are unique instances of the Metropolis algorithm [71]. The former integrates Hamiltonian dynamics with gradient information and auxiliary mass functions to sample from the joint posteriors, whereas the latter derives the proposal distribution from one iteration of the Iterative Weighted Least Squares algorithm [IWL; 72] and generates pseudo-responses using the transformation [73, p. 116] and proceeds iteratively from there on. Although these procedures work in practice, they do not directly capitalize on the conjugacy of the parameters as one would with conducive likelihoods.

The Polya-Gamma latent configuration [68] addresses the above concerns. This missing data augmentation is trivial to construct and can accommodate fairly complicated models, resulting in feasible computational run time. Integrating different hierarchical and nested priors into a logistic conditional model is mathematically more convenient and direct. This is because the Polya-Gamma scheme provides an exact alternative form to the binomial likelihoods that is more tractable and separable, allowing for easy derivation of joint posteriors and making the most of potential conjugate structures. Although there are other missing-data strategies for the logistic model [74]–[76], these involve data augmentation procedures that are either approximate or fairly complicated as they involve multiple layers of latent variables.

Some work has been done in extending the Polya-Gamma data augmentation to models with an autoregressive structure. Pillow and Scott [77] used the Polya-Gamma in factor analysis for negative-binomial spiking with a vector autoregressive structure for the latent factors. Kook, Vaughn, DeMaster, *et al.* [78] who combined a Polya-Gamma latent configuration with an autoregressive structure and Variational Inference. Although this model

does have an autoregressive latent structure, it does not integrate random effects as per conventional GLMM protocols. Further, the model does not allow for partial autocorrelations (Section 1.3.1). Koki, Meligkotsidou, and Vrontos [79] modeled a time series data set using a Polya-Gamma data-augmentation for an AR(k) process. This model is not applicable to our data since the autoregressive structure in [79] manifests itself through the linear predictors, whereas a repeated measures model uses a covariance structure to directly model this, while controlling for predictors. Krisztin and Piribauer [80] extended the Polya-Gamma data augmentation to a spatial AR(1) model. However, this model is lacking for our purposes in that it does not allow for an AR(k) structure.

This work extends the Polya-Gamma latent configuration for modelling AR(k) temporal structures of binary outcomes. This is done by placing an uninformative prior on the covariance structure that can specifically accommodate a flexible AR(k) process with added heterogeneity and other random effects. As a result, we can model correlated binary outcomes with large number of repeat measurements that may have nested AR(k) structures. We refer to this new model as the General Logistic Autoregressive Model With Random Effects (GLAMRE).

This chapter is organized as follows: Section 2.2 will provide formal notation and assumptions on a GLMM for a binary repeat measurements model with random effects, a review on the BGLIMM procedure in SAS® and the Polya-Gamma random variable. Section 2.3 will introduce our GLAMRE model, detailing the derivation of the Gibbs sampler and Metropolis-Hastings algorithm for our procedure. Section 2.4 will compare our GLAMRE model with PROC BGLIMM and PROC GLIMMIX for varying AR(k) settings. Section 2.5 will demonstrate the GLAMRE model on the National Institute of Health (NIH) study from Section 1.5 that measures the effects of service dogs on the daily lives of post-9/11 military Veterans diagnosed with post-traumatic stress disorder (PTSD). Finally, Section 2.6 will include concluding remarks on the GLAMRE model and possible extensions for future research opportunities.

2.2 Background

2.2.1 Notation

For subject $\mathrm{i} = 1, \ldots, I$ with $\mathrm{j} = 1, \ldots n_\mathrm{i}$ repeat measurements and $c = 1, \ldots C$ different cohorts, the GLMM for a binary repeated measurements model with time and cohort random effects is defined as

$$\underset{n_\mathrm{i}\times 1}{\boldsymbol{\eta}_\mathrm{i}} = \underset{n_\mathrm{i}\times p}{\boldsymbol{X}_\mathrm{i}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n_\mathrm{i}\times(n_\mathrm{i}+C)}{\boldsymbol{Z}_\mathrm{i}}\ \underset{(n_\mathrm{i}+C)\times 1}{\boldsymbol{\gamma}_\mathrm{i}}, \text{ where } \boldsymbol{\gamma}_\mathrm{i} \sim N(\mathbf{0}, \boldsymbol{G}) \tag{2.1}$$

with the accompanying link function $\mathbb{E}\Big[\boldsymbol{Y}_\mathrm{i}|\boldsymbol{\beta}, \boldsymbol{\gamma}\Big] = \boldsymbol{\mu}_\mathrm{i} = g^{-1}(\boldsymbol{\eta}_\mathrm{i}) = \frac{1}{1+\mathrm{e}^{-\boldsymbol{\eta}_\mathrm{i}}}$. An alternative representation of this would be

$$\underset{n_\mathrm{i}\times 1}{\boldsymbol{\eta}_\mathrm{i}} = \underset{n_\mathrm{i}\times p}{\boldsymbol{X}_\mathrm{i}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n_\mathrm{i}\times n_\mathrm{i}}{\boldsymbol{Z}_\mathrm{time}}\underset{n_\mathrm{i}\times 1}{\boldsymbol{\gamma}_\mathrm{time,i}} + \underset{n_\mathrm{i}\times C}{\boldsymbol{Z}_\mathrm{cohort,i}}\underset{C\times 1}{\boldsymbol{\gamma}_\mathrm{cohort}} \tag{2.2}$$

Here $\boldsymbol{X}_\mathrm{i}$ is the design matrix for the fixed effects for subject i, where each column represents a predictor, beginning with a column of 1's for the intercept term. $\boldsymbol{Z}_\mathrm{i}$ is the design matrix for the random effects for subject i. Specifically, $\boldsymbol{Z}_\mathrm{time}$ is an identity matrix of size n_i, i.e. $\mathbb{1}_{n_\mathrm{i}\times n_\mathrm{i}}$ while $\boldsymbol{Z}_\mathrm{cohort,i}$ is a $n_\mathrm{i} \times C$ matrix that has a column of 1's in the c-th column and 0 everywhere else if subject i is from cohort c. The $\boldsymbol{\beta}$ matrix contains the fixed effects while the $\boldsymbol{\gamma}$ matrix contains the random effects. In a GLMM paradigm, the random effects $\boldsymbol{\gamma}$ is assumed to be normally distributed with mean $\mathbf{0}$ and some covariance matrix $\boldsymbol{G}$. For a comprehensive list of possible covariance structures for $\boldsymbol{G}$, readers can refer to the SAS® manual [14]. Finally, the link function is the continuous $g(.)$ function that connects our observed response to the data.

2.2.2 Literature Review

PROC BGLIMM in SAS® is a simulation-based procedure that draws inference based on the joint posterior distribution of parameters in a GLMM model. More informally, it is the Bayesian version of PROC GLIMMIX. PROC BGLIMM uses the Gibbs sampler by default to update conditional draws in which the fixed parameters are sampled jointly at

each iteration. Contingent upon the user's request, the random effects can be updated either jointly or by clusters.

This procedure is equipped with the Hamiltonian Monte Carlo (HMC) sampler to draw from the joint posteriors of the parameters. HMC is a specialized version of the Metropolis algorithm that uses Hamiltonian dynamics supplemented with gradient information and auxiliary momentum variables to sample from the target distribution [69]. Specifically, HMC combines the target distribution or posterior of the model parameters $\boldsymbol{\pi}(\boldsymbol{\theta})$ with the auxiliary momentum variable $\boldsymbol{r}$ into the following potential energy function $\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{r}) \propto \boldsymbol{\pi}(\boldsymbol{\theta}) \exp\left\{-\frac{1}{2}\boldsymbol{r}^T\boldsymbol{r}\right\}$.

The HMC then samples from the joint space of $(\boldsymbol{\theta}, \boldsymbol{r})$ by first sampling $\boldsymbol{r}$ from a standard normal, then discarding the $\boldsymbol{r}$ draws and retaining the sampled $\boldsymbol{\theta}$ as samples from $\boldsymbol{\pi}(\boldsymbol{\theta})$. This is achieved by moving along the gradient trajectory, typically using the leapfrog method with tuning parameters step size ϵ and L steps. This is repeated for L times. Finally, the proposed values $(\boldsymbol{\theta}^*, \boldsymbol{r}^*)$ are accepted with probability $\min\left\{1, \frac{\boldsymbol{\pi}(\boldsymbol{\theta}^*, \boldsymbol{r}^*)}{\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{r})}\right\}$.

PROC BGLIMM has a built-in adaptive HMC which uses the No U-Turn Sampler [NUTS; 81] to automatically tune for ϵ and L based on some supplied target acceptance probability δ. The NUTS algorithm does this by building a binary tree around leaf nodes that represent the states of $(\boldsymbol{\theta}, \boldsymbol{r})$. At each state, the tree branches left or right and takes 2^{j} leapfrog steps of size ϵ, where j is the current height of the binary tree. This binary expansion continues until a sampling particle makes a U-Turn and revisits a state it has once explored. This is process is repeated until the ϵ is tuned such that the acceptance rate in this branching process is close to the supplied value δ.

A Polya-Gamma random variable X with parameters $b > 0$ and $c \in \mathbb{R}$ is defined as

$$X \overset{D}{=} \frac{1}{2\pi^2} \sum_{\mathrm{i}=1}^{\infty} \frac{g_k}{\left(k - \frac{1}{2}\right)^2 + \frac{c^2}{4\pi^2}}, \text{ where } g_k \overset{\mathrm{iid}}{\sim} Ga(b, 1) \tag{2.3}$$

The advantage of the Polya-Gamma random variable is that the binomial likelihoods used in logistic models can now be represented as mixtures of Gaussians with respect to a Polya-Gamma distribution [68], i.e.

$$\frac{(\mathrm{e}^{\psi})^a}{(1+\mathrm{e}^{\phi})^b} = 2^{-b}\mathrm{e}^{k\psi}\int \mathrm{e}^{\frac{-\omega\psi^2}{2}}p(\omega)d\omega \tag{2.4}$$

where $k = a - \frac{b}{2}$ and $\omega \sim PG(b, 0)$. Note that when $\psi = x^T\boldsymbol{\beta}$, the integrand on the right is the kernel of a Gaussian likelihood in $\boldsymbol{\beta}$. This result yields a simple Gibbs strategy to sampling the posteriors from binomial likelihoods: Gaussian draws for $\boldsymbol{\beta}$ alternated with Polya-Gamma draws for the single layer of latent variables.

2.3 Methodology

2.3.1 Priors

We assume a normal prior for the fixed effects $\boldsymbol{\beta}$ such that $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta_0}, \boldsymbol{\Sigma}_\beta)$. Specifically, we assume a flat prior for the fixed effects $\boldsymbol{\beta}$ using a Normal distribution with mean and precision $\mathbf{0}$ [4].

For the k-partial autocorrelations $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_k)$ (Section 1.3.1), we assign a flat prior over the k-cuboid parameter space: $\boldsymbol{\rho} \sim \text{Uniform}_k(-1, 1)$. Although this flat prior is convenient, the parameter space is constrained on the (-1,1) space. This makes sampling from the posterior of $\boldsymbol{\rho}$ inconvenient since the target distribution would have to be truncated based on the prior boundary space. To circumvent this, we apply the Fisher transformation (or atanh()) on the partial autocorrelations $\boldsymbol{\rho}$ [31]. As a result, we now have the following prior $\rho_k^z = \text{atanh}(\rho_k) \underset{\text{iid}}{\sim} \text{logistic}\Big(\mu = 0, s = \frac{1}{2}\Big)$.

We assign an inverse gamma prior to both the variance from time and cohort with v_0 for both the shape and scale parameters, i.e. $\sigma^2_{\text{time}} \sim IG(v_0, v_0)$ and $\sigma^2_{\text{cohort}} \sim IG(v_0, v_0)$. We set $v_0 = 1$ to reflect a non-informative prior with conditional conjugacy [82]. The autoregressive structure manifests itself through the random effect $\boldsymbol{\gamma}_{\text{time}}$, i.e. $\boldsymbol{\gamma}_{\text{time,i}}|\boldsymbol{\rho}_k^z, \sigma^2_{\text{time}} \sim N_{n_i}(\mathbf{0}, \sigma^2_{\text{time}}\boldsymbol{G})$, where $\boldsymbol{G} = r(\tanh(\boldsymbol{\rho}_k^z))$. Here $r(.)$ is the recurrence function that deploys the partial autocorrelations into the correlation matrix from [30]. Similarly, the random effect

from cohort manifests itself through the variable $\gamma_{\text{cohort}}|\sigma^2_{\text{cohort}} \sim N(0, \sigma^2_{\text{cohort}})$. Finally, we have the latent configuration from the Polya-Gamma random variable $\boldsymbol{\omega}_{\text{ij}} \sim \text{PG}(n_{\text{ij}} = 1, 0)$, for binary outcomes.

To summarize, our sampler would be drawing from the posteriors of the parameters $\boldsymbol{\theta}$ at each iteration with p predictors (including intercept), k partial autocorrelations, 2 independent sources of variability from both time and cohort, IJ random effects from time, C cohort random effects as well as IJ Polya-Gamma latent variable.

2.3.2 Posterior

We will exploit the notation $\boldsymbol{\theta}$ to conveniently represent the remaining parameters in a joint posterior distribution. Following the prior configuration with the Polya-Gamma latent setup, the posterior of the fixed effect is given by

$$\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{y} \sim N_p\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right) \tag{2.5}$$

where $\boldsymbol{\Sigma}_\beta = \left(\boldsymbol{\Sigma}_b^{-1} + \sum_{\text{i}=1}^N \boldsymbol{X}_{\text{i}}^T \boldsymbol{\Omega}_{\text{i}} \boldsymbol{X}_{\text{i}}\right)^{-1}$ and $\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta \left(\boldsymbol{\Sigma}_b^{-1}\boldsymbol{b} + \sum_{\text{i}=1}^N \boldsymbol{X}_{\text{i}}^T \boldsymbol{\Omega}_{\text{i}} \boldsymbol{l}_{\text{i}}\right)$. Additionally, $\boldsymbol{l}_{\text{i}} = (\frac{k_{\text{i}1}}{\omega_{\text{i}1}} - (\boldsymbol{z}_{\text{time},1}\boldsymbol{\gamma}_{\text{time,i}} + \boldsymbol{z}_{\text{cohort,i}}\boldsymbol{\gamma}_{\text{cohort}}), \ldots, \frac{k_{\text{i}J}}{\omega_{\text{i}J}} - (\boldsymbol{z}_{\text{time},n_{\text{i}}}\boldsymbol{\gamma}_{\text{time,i}} + \boldsymbol{z}_{\text{cohort,i}}\boldsymbol{\gamma}_{\text{cohort}}))^T$ and $\boldsymbol{\Omega}_{\text{i}} = diag(\omega_{\text{i}1}, \ldots, \omega_{\text{i}J})$.

Similarly, the posterior random effects from the random factor time that manifest from σ^2_{time} is given by

$$\boldsymbol{\gamma}_{\text{time,i}}|\boldsymbol{\theta}, \boldsymbol{y}_{\text{i}} \sim N_{n_{\text{i}}}(\boldsymbol{\mu}_{\gamma_{\text{time,i}}}, \boldsymbol{\Sigma}_{\gamma_{\text{time,i}}}) \tag{2.6}$$

where $\boldsymbol{\Sigma}_{\gamma_{\text{time,i}}} = \left((\sigma^2_{\text{time}}\boldsymbol{G})^{-1} + \boldsymbol{Z}^T_{\text{time,i}}\Omega_{\text{i}}\boldsymbol{Z}_{\text{time,i}}\right)^{-1}$ and $\boldsymbol{\mu}_{\gamma_{\text{time,i}}} = \boldsymbol{\Sigma}_{\gamma_{\text{time,i}}}\boldsymbol{Z}^T_{\text{time,i}}\Omega_{\text{i}}\tilde{\boldsymbol{l}}_{\text{i}}$. Subsequently, $\tilde{\boldsymbol{l}}_{\text{i}} = (\frac{k_{\text{i}1}}{\omega_{\text{i}1}} - (\boldsymbol{x}_{\text{i}1}\boldsymbol{\beta} + \boldsymbol{z}_{\text{cohort,i}}\boldsymbol{\gamma}_{\text{cohort}}), \ldots, \frac{k_{\text{i}J}}{\omega_{\text{i}J}} - (\boldsymbol{x}_{\text{i}J}\boldsymbol{\beta} + \boldsymbol{z}_{\text{cohort,i}}\boldsymbol{\gamma}_{\text{cohort}}))^T$. The posterior random effects from the random factor cohort that manifest from σ^2_{cohort} is similarly given by

$$\boldsymbol{\gamma}_{\text{cohort}}|\boldsymbol{\theta}, \boldsymbol{y} \sim N_C\left(\boldsymbol{\mu}_{\text{cohort}}, \boldsymbol{\Sigma}_{\text{cohort}}\right) \tag{2.7}$$

where $\boldsymbol{\Sigma}_{\text{cohort}} = \left(\frac{1}{\sigma^2_{\text{cohort}}}\mathbb{1}_{C\times C} + \sum_{\text{i}=1}^{I} \boldsymbol{Z}^T_{\text{cohort,i}}\boldsymbol{\Omega}_{\text{i}}\boldsymbol{Z}_{\text{cohort,i}}\right)^{-1}$ and $\boldsymbol{\mu}_{\gamma_{\text{cohort}}} = \boldsymbol{\Sigma}_{\text{cohort}} \sum_{\text{i}=1}^{I} \boldsymbol{Z}^T_{\text{cohort,i}}\boldsymbol{\Omega}_{\text{i}}\breve{\boldsymbol{l}}_{\text{i}}$. Additionally, $\breve{\boldsymbol{l}}_{\text{i}} = \left(\frac{k_{\text{i}1}}{\omega_{\text{i}1}} - (\boldsymbol{x}_{\text{i}1}\boldsymbol{\beta} + \boldsymbol{z}_{\text{time,j}}\boldsymbol{\gamma}_{\text{time,i}}), \ldots, \frac{k_{\text{i}J}}{\omega_{\text{i}J}} - (\boldsymbol{x}_{\text{i}J}\boldsymbol{\beta} + \boldsymbol{z}_{\text{time},n_{\text{i}}}\boldsymbol{\gamma}_{\text{time,i}})\right)^T$.

The posteriors for the variance from time σ^2_{time} and the cohort σ^2_{cohort} are given by

$$\sigma^2_{\text{time}}|\boldsymbol{\theta}, \boldsymbol{y} \sim IG\left(\left(\frac{IJ}{2} + v_0\right), \frac{2v_0 + \sum_{\text{i}=1}^{I} \boldsymbol{\gamma}^T_{\text{time,i}}\boldsymbol{G}^{-1}\boldsymbol{\gamma}_{\text{time,i}}}{2}\right) \tag{2.8}$$

$$\sigma^2_{\gamma_{\text{cohort}}}|\boldsymbol{\theta}, \boldsymbol{y} \sim IG\left(\left(\frac{C}{2} + v_0\right), \frac{2v_0 + \boldsymbol{\gamma}^T_{\text{cohort}}\boldsymbol{\gamma}_{\text{cohort}}}{2}\right) \tag{2.9}$$

Based on Polson, Scott, and Windle [68], the posterior of the Polya-Gamma random variable is

$$\omega_{\text{ij}}|\boldsymbol{\theta}, \boldsymbol{y} \sim PG(n_{\text{ij}} = 1, \boldsymbol{x}^T_{\text{ij}}\boldsymbol{\beta} + \boldsymbol{z}^T_{\text{ij}}\boldsymbol{\gamma}_{\text{i}}) \tag{2.10}$$

The posterior of $\boldsymbol{\rho}^z$ is given by

$$\boldsymbol{\pi}(\boldsymbol{\rho}^z|\boldsymbol{\theta}, \boldsymbol{y}) \propto \left[\prod_{\text{i}=1}^{I} N_{n_{\text{i}}}\left(\boldsymbol{\gamma}_{\text{time,i}}|\boldsymbol{0}, \sigma^2_{\text{time}}\boldsymbol{G}\right)\right]\left[\prod_{k=1}^{K} \text{logistic}\left(\rho_k|\mu = 0, s = \frac{1}{2}\right)\right] \tag{2.11}$$

$$\boldsymbol{G} = r(\tanh(\boldsymbol{\rho}^z_k)), r(.) = \text{ recurrence relation from [30]}$$

We will employ Metropolis-Hastings to sample from the posteriors of the $\boldsymbol{\rho}^z$, which are the transformed partial autocorrelations, using the symmetric proposal $N_K\left(\boldsymbol{\rho}^{z^{(t)}}|\boldsymbol{\rho}^{z^{(t-1)}}, \sigma^2_{\text{prop}}\boldsymbol{I}_K\right)$. Through simulations and experiments, a proposal value of $\sigma^2_{\text{prop}} = 0.01$ was selected since it explored the partial autocorrelation space most effectively. The above Polya-Gamma sampler is summarized in Figure 2.1. All derivations are provided in Section B.B.

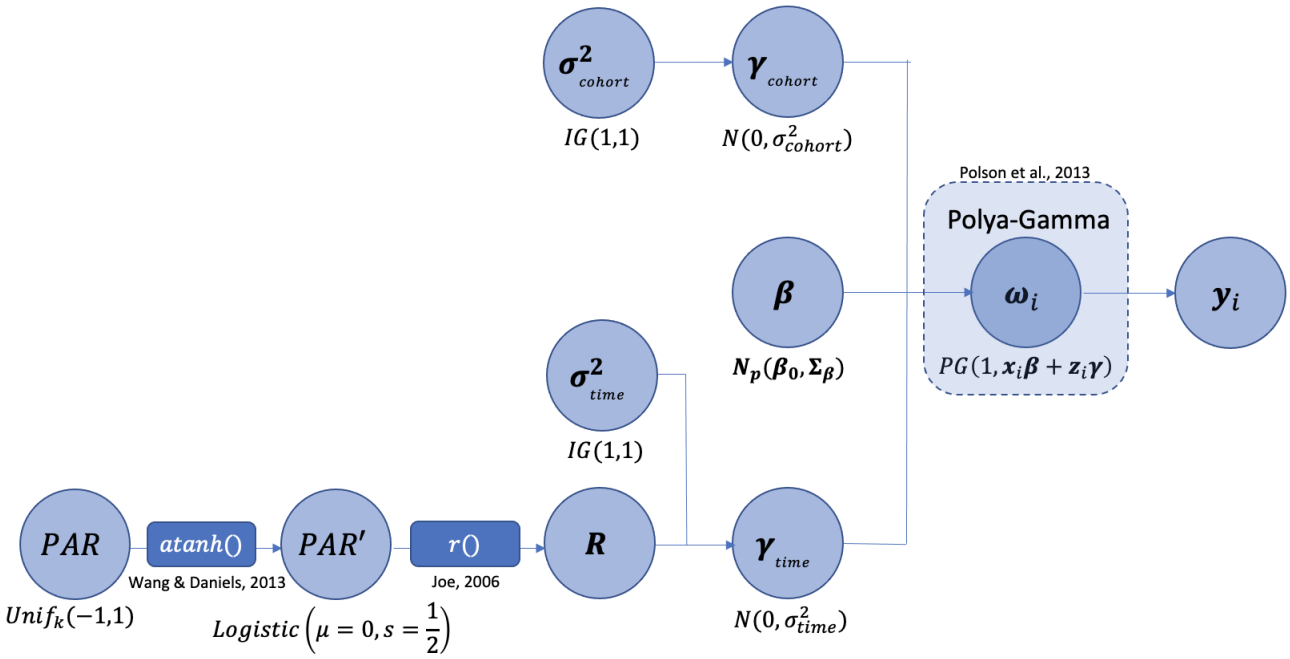


Figure 2.1. Summary of the GLAMRE procedure.

2.4 Simulation

2.4.1 AR(1) Simulation

In the AR(1) simulation study, we generated a total of 1000 data replicates for 150 subjects, each with 30 repeat measurements. There are 25 levels for cohort, with each cohort having 6 subjects. Each simulated subject would have a continuous covariate sampled from a standardized normal distribution. There are three estimable fixed effects $(\beta_0, \beta_{\text{trt}}, \beta_1)$, where β_0 is the intercept and β_1 is the fixed effect associated with the generated continuous covariate, two sources of random variability $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}})$ and a scalar partial autocorrelation ρ. In applying an intention-to-treat analysis on the preliminary data, we fix (β_0, β_1) = (-2, 1.5), and the β_{trt} parameter setting is chosen from $(0, 0.25, 0.5)$. In addition, the random sources of variability were set to $\sigma^2_{\text{time}} = (2, 20)$ and $\sigma^2_{\text{cohort}} = (0.5, 5)$ respectively. Finally, the partial autocorrelation for the AR(1) setting was set to $(-0.7, -0.5, -0.3, 0.3, 0.5, 0.7)$. We compared the performance of our model against BGLIMM procedure in SAS® using HMC with the NUTs configuration. The AR(1) simulation settings are summarized in Table 2.1. We assess model performance using power through posterior predictive p-values [33] from

Section 1.3.3, bias, coverage and standard error over the 1000 data replicates. Code for the BGLIMM procedure is provided in Appendix B.C.

Table 2.1. Input parameters for AR(1) simulations with 150 subjects and 30 repeated measurements each.

β_{trt}	AR(1)	σ^2_{time}	σ^2_{cohort}	Method
0	0.7	2	0.5	GLAMRE
0.25	0.5	20	5	BGLIMM
0.5	0.3			
	-0.3			
	-0.5			
	-0.7			

GLAMRE has generally higher power and better treatment coverage than BGLIMM (Figures B.1-B.4). Although BGLIMM does exhibit higher power in Figure B.2, this difference is somewhat small. Whereas we can see fairly clearly that GLAMRE has a distinctly higher power in Figure B.3. The coverage for GLAMRE and BGLIMM hover around 0.95, with the exception of when there is high variability in time and cohort (Figure B.4) wherein we see the treatment coverage for BGLIMM dipping lower than 0.9. GLAMRE has lower bias when variability due to time is higher but higher bias when variability due to both time and cohort are lower (Figure B.1).

GLAMRE has generally better coverage and lower bias than BGLIMM for the PARs (Figures B.5-B.8). Although GLAMRE has a larger bias and standard error in recovering σ^2_{time} and σ^2_{cohort}, these inflated values compensate for its high coverage. In comparison, BGLIMM has much lower coverage for both sources of variability when $\sigma^2_{\text{time}} = 20$ and $\sigma^2_{\text{cohort}} = 5$ (Figures B.9-B.16). Since GLAMRE captures the variability better, this translates to a higher standard error in the treatment posteriors.

2.4.2 AR(2) Simulation

The AR(2) study has similar settings in the AR(1) study, just that now the AR(2) parameters are set to 8 different settings: $(-0.7, -0.5), (-0.7, 0.5), (0.7, -0.5), (0.7, 0.5)$, $(-0.5, -0.3), (-0.5, 0.3), (0.5, -0.3), (0.5, 0.3)$. As before, model performances are evaluated based on power, bias, coverage and standard error. We compare the performances of GLAMRE with AR(1), AR(2), and AR(3) settings with BGLIMM which is limited to an AR(1) setting. Model fit will be assessed using the Deviance Information Criterion (DIC) and further supplemented by the 95% credible intervals for the PAR posteriors. Simulation settings for the AR(2) study is summarized in Table 2.2.

Table 2.2. Input parameters for AR(2) simulations with 150 subjects and 30 repeated measurements each.

β_{trt}	AR(2)	σ^2_{time}	σ^2_{cohort}	Method
0	(-0.7, -0.5)	2	0.5	GLAMRE AR(1)
0.25	(-0.7, 0.5)	20	5	GLAMRE AR(2)
0.5	(0.7, -0.5)			GLAMRE AR(3)
	(0.7, 0.5)			BGLIMM
	(-0.5, -0.3)			
	(-0.5, 0.3)			
	(0.5, -0.3)			
	(0.5, 0.3)			

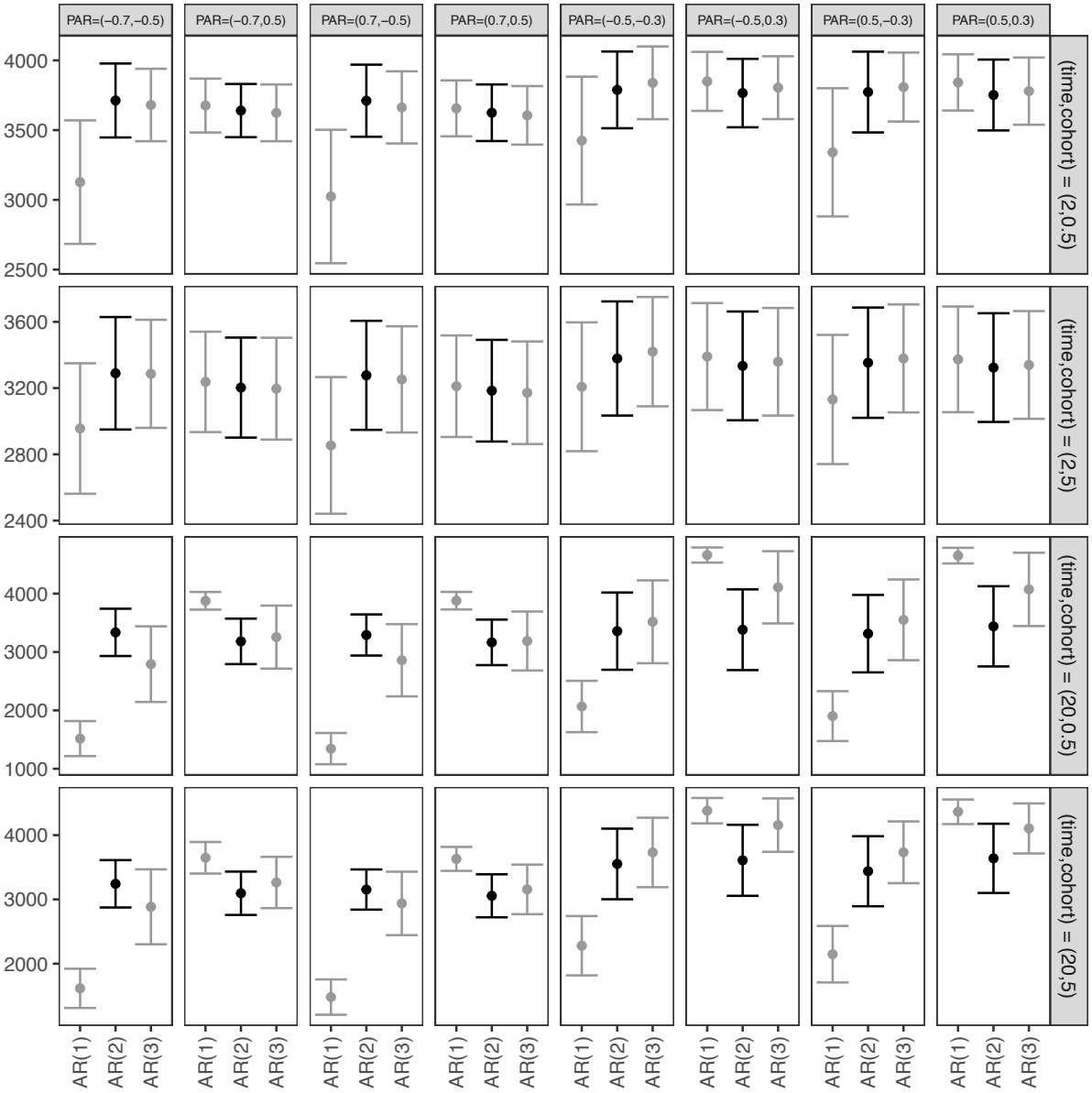
Similar to the results from the AR(1) study, the correctly specified GLAMRE-AR(2) has better, if not similar power than the rest (Figures B.17-B.20). As we expect, treatment estimates from the incorrectly specified GLAMRE-AR(1) is overly biased and suffers from undercoverage. This is especially prominent when time and cohort have high values of variability (Figure B.20). Similarly, BGLIMM has coverage that dips lower than 0.9 when there is higher variability associated with time (Figures B.19 and B.20).

For the random effects, we can see that the GLAMRE-AR(2) does a good job of recovering the true PARs (Figure B.21-B.24). In addition, we note that although GLAMRE-AR(3)

is a misspecified model, the last PAR posterior is often close to zero, yielding a marginal correlation that is effectively an AR(2) structure. Similarly, GLAMRE-AR(2) has consistently low bias and better coverage than all the rest, albeit a slightly higher SE relative to the misspecified GLAMRE-AR(1).

The most appropriate model is selected using the a combination of the DIC fit measure and the CI of the last PAR of the relevant model. From Figure 2.2, we can see that the GLAMRE-AR(2) model is selected correctly when AR(2) = {(-0.7, 0.5), (0.7, 0.7), (-0.5, -.3), (0.5, 0.3)}, while GLAMRE-AR(1) has the lowest DIC measure for the remaining AR(2) settings. One reason for this incorrect model selection is that the corresponding standard errors for $(\beta_{\text{trt}}, \sigma^2_{\text{time}}, \sigma^2_{cohort})$ are consistently inflated over all simulation settings relative to the other models (Figure B.17-B.32). In addition, we have the CIs for the last PARs of the corresponding AR(k) setting in our toolkit to supplement model selection. In all AR(2) simulation settings, 0 is consistently not in the 95% CI of the PAR (Figure B.21-B.24, top row). Subsequently, 95% CI for the last PAR of AR(3) would often cover 0 (Figure B.21-B.24, bottom row). Favoring parsimony, the AR(2) would be selected. Further, we have seen that choosing between AR(2) and AR(3) is less detrimental to model performance than having to choose between AR(1) and AR(2).

Figure 2.2. Plots of DIC measures with their accompanying standard errors over 1000 data replicates when $\beta_{\text{trt}} = 0$ with varying σ^2_{cohort} and σ^2_{time}. Results are the same for other β_{trt} values. While the DIC does select the correct model at times (solid black), it occasionally favors the simple AR(1) model. This can be attributed to inflated standard errors in the estimation of the other parameters in the corresponding model.



2.5 Application: PTSD Clinical Trial

2.5.1 Description Of Study

We demonstrate an application of the GLAMRE methodology on an NIH-funded clinical trial (#NCT03245814) conducted by Dr. Marguerite O'Haire on the effectiveness of service dogs for United States military Veterans diagnosed with post traumatic stress disorder (PTSD) (Section 1.5.1). The data comprises 142 Veterans, each with approximately 28 repeated measurements during the follow-up period.

The goal in analyzing this repeated measures study is to determine the effects of a service dog on the probability that a Veteran is around other people, while accounting for baseline covariates and time dependencies. Vincent, Auger, Lavoie, *et al.* [41] and Crowe, Sanchez, Howard, *et al.* [42] have suggested that pairing a Veteran with a service dog for PTSD would encourage and enable the Veteran to be more involved in their communities and more prone to engage with others. However, the bulk of these results are qualitative in nature. This present NIH study and GLAMRE model add to the current literature by quantitatively analyzing the odds of a Veteran being around other people with the service dog in comparison to a waitlist control group. Results from this analysis can help mental health practitioners and service dog providers by providing additional insight into interventions involving service dogs.

A reasonable covariance structure in this context would use the AR(k) process scaled by some constant σ^2_{time}, which has the correlation in the outcomes between any two points diminishing as a function of their difference in discrete time. In specifying the AR process, we note as the data are collected twice daily, a subject may be less likely to be surrounded by other people with their service dog had they done so the preceding the current time point, but more likely to do so immediately after the current time point.

2.5.2 Model

The outcome is a dichotomous 0-1 variable: the outcome is 1 if the Veteran responded yes to being with children, other family members, friends, acquaintances, coworkers, strangers,

spouse or significant other and 0 otherwise. We are modelling the odds of the Veteran being around other people. Three predictors are used in our model: treatment assignment (service dog versus waitlist group), age of Veterans and the proportion of time the Veterans were with other people at baseline. Age was included as a demographic variable that may related to Veterans being around other people.

There are two separate components that go into the correlation structure of our model: the vector of partial autocorrelations $\boldsymbol{\rho}_{\text{weekday}}$ that govern the AR(k) structure for weekday, defined as Monday through Thursday, and the vector of partial autocorrelations $\boldsymbol{\rho}_{\text{weekend}}$ that govern the AR(k) structure for weekend, defined as Friday through Sunday. This specification is motivated by the weekend effect, which suggests that human behavior is rhythmic and cycles between weekdays and weekends [43]–[46]. To supplement this, we follow the partition of weekends and weekends from Ryan, Bernstein, and Brown [47] that clusters Friday evening through Sunday morning as weekends and Monday through Thursday as Weekdays. We extend this definition of weekend to include Friday morning and Sunday evening since the original weekend partition definition would greatly limit the number of observations to compute the AR(k) structures.

We run the GLAMRE model on three different AR(k) settings for 10000 iterations with a burn-in of 5000: (1) AR(1) for both weekday and weekend, (2) AR(2) for both weekday and weekend and (3) AR(3) for both weekday and weekend.

2.5.3 Results

Table 2.3 summarizes the DIC as well as the treatment posterior summaries for the different AR(k) settings for GLAMRE and AR(1) BGLIMM. GLAMRE with an AR(2) setting for the weekday and weekend partitions were selected using DIC as a the measure. The sensitivity analysis shows the PPP and treatment posterior mean stays stable across the varying AR(k) settings for GLAMRE and BGLIMM. Table B.1 contains a comprehensive list of all random and fixed effects for all the models.

Table 2.4 summarizes the posterior summaries of all the fixed effects in the AR(2) model that was selected. We conclude that although treatment assignment has a positive impact

on the odds of a Veteran being around other people, it is not significant at $\alpha = 0.05$. It is noteworthy that the 95% credible interval for the baseline mean predictor does not cover the zero value. This implies that subject behavior at baseline may have an association with their behavior at followup, i.e. the proportion of time Veterans are around other people at baseline is associated with the odds of them being around people at followup.

Table 2.3. Posterior summaries of treatment effect with varying AR(k) settings for GLAMRE and AR(1) for BGLIMM, with standard deviations within the parentheses. The AR(k) setting for GLAMRE was chosen based on DIC. Treatment effect remains insignificant throughout all settings at α=0.05.

Method	AR(k)	DIC	Treatment Posterior Mean	95% Lower CI	95% Upper CI	PPP
GLAMRE	1	4048	0.19 (0.19)	-0.17	0.57	0.8
	2	3755	0.13 (0.28)	-0.43	0.64	0.81
	3	4058	0.14 (0.20)	-0.22	0.55	0.75
BGLIMM	1	4003	0.09 (0.24)	-0.33	0.61	0.81

Table 2.4. Posterior Summaries of parameters for the AR(2) setting for both weekdays and weekends.

Variable	Posterior Mean	95% Lower CI	95% Upper CI
Intercept	0.62 (0.24)	0.19	1.13
Treatment	0.13 (0.28)	-0.43	0.64
Age	0.001 (0.09)	-0.17	0.16
Baseline Mean	1.54 (0.28)	1.14	2.24

2.6 Discussion

In applying the GLAMRE model to the NIH study, we conclude that although there is a positive association between being assigned a service dog and being around other people at the 3-months followup period, this effect is not significant. Rather, there is a larger association between behavior at baseline and follow-up on the participant level, such that

Veterans who were more likely to be around people were more likely to do so at 3-months follow-up. This suggests that there may be robust individual differences in daily routines and propensity to be around other people that are stronger than otherwise measured in this study.

Regardless, these findings do not paint a comprehensive picture of how service dogs can affect Veterans with PTSD. Although prior studies suggest that service dogs may assist Veterans in being more involved with their communities and being more comfortable being in public settings with other people [48]–[51], this may not have been represented consistently enough in this NIH study for the signal to be detected. Specifically, this qualitatively reported effect could be entangled with significant confounders not recorded in this quantitative analysis. Further, the outcome survey is inquiring whether the Veteran was with other people at the specific time. It could be that the Veteran was with other people during other times of the day, but was alone during the times they responded to the questionnaire. In addition, being with other people is not the end all be all measure that fully defines the efficacy of a service dog in helping Veterans with PTSD.

Although assuming a unstructured covariance matrix would have been a more general assumption than an AR(k) process, the number of repeat measures in this study warrants a longer than average computational power. For example, running the GLIMMIX procedure in SAS® on a single simulated data from Section 2.4 did not converge even after 15 hours of runtime.

A key feature of the GLAMRE model is that it is a Bayesian Model that can used directly in causal models such as Principal Stratification [53] to infer conditional causal effects in longitudinal studies. Although Frangakis, Brookmeyer, Varadhan, *et al.* [54] and Wang, Jo, and Hendricks Brown [55] have implemented these causal models for longitudinal studies, a flexible and nested autoregressive structure has never been used with random effects and heterogeneity in the correlation matrix, as of the submission of this paper. This would be fairly useful in disentangling causality in drug interventions with non-compliance in longitudinal studies [56].

In conclusion, the GLAMRE is a Bayesian Model that allows users to have varying and heterogeneous nested AR(k) structures in the same model with random effects, all inter-

pretable on the log odds scale equipped with high power for binary outcomes in longitudinal studies.

3. CITIES: CLINICAL TRIALS WITH INTERCURRENT EVENTS SIMULATOR

3.1 Introduction

Randomized controlled clinical trials (RCTs) begin their lives as designed experiments that control for baseline covariates to assess the effect of a treatment or intervention [83]. Throughout the course of the trial, patients may inevitably discontinue their randomized assigned treatment due to lack of efficacy (LOE), excess efficacy (EE), adverse effects (AE) or administrative reasons [84]. These disruptions to the planned clinical trial protocol can muddle or confound the true treatment response of both experimental and control treatments being studied [85]. In the language of the recent International Council of Harmonization (ICH) Guidelines on this matter [86], such disruptions are labelled intercurrent events (ICEs). The focus of this work is on modelling and simulating clinical trials that incorporate realistic scenarios for the discontinuation of study treatments in randomized, controlled clinical trials. This is a very pervasive and important problem in pharmaceutical drug development and indeed in academic and governmental funded clinical trials as well.

More formally, ICEs are "events that occur after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest". For instance, terminal events such as death can be seen as an extreme AE [87] that would affect the existence rather than missingness of a measurement. Clinical events can also be attributed to ICEs: subjects who experience excessive reduction in their Haemoglobin A1c (Hb1Ac) levels from a diabetes medication may form a sub-population that is distinct from the target population of interest, i.e. those who experience hypoglycemia versus those who were able to adhere to their treatment protocol. These effects are especially amplified in clinical trials with repeat measurements, since there are more chances of ICEs to happen. Simulated clinical trials should also capture this reality to reflect a more holistic setting.

Discontinuation in clinical trials and other ICEs bring into question the validity and utility of traditional methods such as intent-to-treat (ITT) into question, where analysis is carried out based on the planned treatment regimen. Consider a trial with two treatment arms. As is common in clinical trials, some patients may discontinue treatment due to adverse reactions to the assigned treatment, some may have lost interest in participation and others may have simply moved to

a new location and are no longer accessible. A conventional model such as ITT would proceed with analysis based on their treatment assignment, disregarding their adherence. This effect is misleading since it does not incorporate how safe the treatment is. Methods that use causal inference have been suggested as a viable path forward to asses the direct treatment effect, without being confounded by the ICEs [88].

Simulators function by generating some reality of ICEs based on the supplied input parameters from the users without users having to know the inner workings of the algorithm. As a consequence, simulators are often viewed as black boxes. Although there are comprehensive softwares such as Facts®, Certara® and Cytel® to simulate clinical trials, these industry products are fairly expensive. Those in academic institutions with limited funding will have to either depend on statistical packages that are free or code the simulators from the ground up. Regardless, these software packages contribute to the perception of a black box simulator, which are not helpful in facilitating meaningful statistical discussions between statisticians, clinicians and medical providers.

Although there are many clinical trials simulators, they do not incorporate ICEs in a causal setting and often warrant prior programming literacy. Sofrygin, Laan, and Neugebauer [89] developed the simcausal R package for specification and simulation of complex longitudinal data structures using Non-Parametric Structural Equations Model (NPSEM) that can be represented using Directed Acyclic Graphs (DAG) [90]. This package allows for correlation that can be integrated using copulas with provided example with syntax that is systematically clean and causal graphs are outputted directly. Unfortunately, the package does not allow users to directly incorporate functions of discontinuation form varying sources of ICEs. Paux and Dmitrieniko [91] developed the mediana R package is a framework for simulating, modelling and evaluating clinical trials with multiple endpoints based on the Clinical Scenario Evaluation (CSE) approach from Benda, Branson, Maurer, *et al.* [92] and further refined in Friede, Nicholas, Stallard, *et al.* [93]. This software package allows users to simulate data from a wealth of distributions such as negative binomial and truncated distributions. However, as before, this package does not provide the flexibility to integrate different functions of discontinuation form varying sources. Further, required programming literacy in using these softwares impedes usage from non-programmers such as medical professional and providers who interact directly with patients, further wedging the disconnect between statisticians and clinicians.

Due to the nature of the pharmaceutical industry, clinical trials data are not easily shareable. This means that novel causal models that have been developed by different companies on their own clinical trials may not be as readily testable on other clinical trials from different companies. This will impede the development of causal models in the pharmaceutical industry, since publications will include only summary information and performance metrics of their models on their own clinical trials that cannot be easily shared with others.

We address the issues highlighted using the Clinical Trials with Intercurrent Events Simulator (CITIES): an Rshiny app written in R for simulating clinical trials with multiple endpoints using the potential outcomes [94]. Although there are several Rshiny apps available for simulating clinical trials [95]–[98], none have been used in simulating potential outcomes and summarizing causal effects in presence of ICEs. The CITIES can incorporate varying sources of discontinuation with varying functional behavior cleanly and directly with graphical representation, demystifying the notion of a black box simulator. Further, users can interact dynamically with the simulator without having to know how to code in R, which makes for a convenient tool for discussing causality in clinical trials. In addition, by having a transparent and intuitive simulator using potential outcomes, researchers are now afforded a simulator that can be used to compare performances of different causal models. Users can also mimic and get quick causal assessments of real clinical trials from publications without having access to the raw data.

CITIES can be seen as a direct application of potential outcomes using the tripartite framework [88], which forms the genesis of this work. In trying to disentangle the pure causal effects of a treatment intervention in presence of ICEs, [88] proposed the tripartite approach - three estimands ('what is to be estimated') that are relevant and meaningful not only to patients, prescribers and payers, but also sponsors and regulators. These tripartite estimands are: (1) probability of discontinuation due to AE, (2) probability of discontinuation due to LOE, and (3) treatment effect in patients who can adhere to the investigational treatment.

We begin by outlining how the potential outcomes are generated, how the varying sources of discontinuation are integrated, and how the causal effects are calculated with the accompanying percentage discontinuation summary in Section 3.2. This is followed by a demonstration of an application of CITIES on two clinical trials in Section 3.3. Finally, concluding remarks on clinical trials simulators with ICE and potential extensions will be provided in Section 3.4.

3.2 Implementation

CITIES is a web application written in Rshiny that can be run from any web browser. The application requires a few manual inputs that are visualized in adjacent panels. CITIES allows for simulation input parameters to be bookmarked and saved for online exploration and sharing. All visuals are updated synchronously and can be interacted with dynamically by hovering the mouse over the visuals. There are four tabs to CITIES that users will navigate through: The Mean Settings tab, the LOE & EE tab, the AE tab and the causal effects tab. To save generated data sets, users will run the Rscript which will save the simulated clinical trials on the local machine.

3.2.1 Mean Settings

Define $y_{\text{ij}t}$ to be the potential outcome for subject $\text{i} = 1, \ldots, I$ at time $\text{j} = 1, \ldots J$ on treatment arm $t = 0$ for control and 1 for test. Assume $y_{\text{i0}t}$ to be the baseline measurement taken at time $t = 0$ that has the same response distribution across all treatment arms before the treatment is assigned. The potential outcomes have a joint Multivariate Normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ follows an autoregressive(k) structure scaled by parameter σ^2.

$$\begin{pmatrix} y_{\text{i0}t} \\ \tilde{\boldsymbol{y}}_{\text{i}t} \end{pmatrix} \sim N \left[\boldsymbol{\mu} = \begin{pmatrix} \mu_{\text{i0}t} \\ \tilde{\boldsymbol{\mu}}_{\text{i}t} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right] \tag{3.1}$$

where $\tilde{\boldsymbol{y}}_{\text{i}t} = \begin{pmatrix} y_{\text{i1}t} & \ldots & y_{\text{i}J t} \end{pmatrix}^T$, $\tilde{\boldsymbol{\mu}}_{\text{i}t} = \begin{pmatrix} \mu_{\text{i1}t} & \ldots & \mu_{\text{i}J t} \end{pmatrix}^T$ and $\boldsymbol{\Sigma} = \boldsymbol{\sigma^2} \times \mathbf{AR(k)}$, $k = 1, \ldots, J-1$. Conditional on the baseline measurement $y_{\text{i0}t}$, CITIES will generate the potential outcomes for each subject across both treatment arms

$$\left(\tilde{\boldsymbol{y}}_{\text{i}t} | y_{\text{i0}t} = y \right) \sim N \left[\tilde{\boldsymbol{\mu}}_{\text{i}t} + \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \quad \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \right]$$

CITIES will then generate trials based on the requested number of data sets, number of subjects per arm, the means for the corresponding treatment arms, the associated scale parameter for the covariance structure and the AR(k) value. We populate the correlation matrix from a vector of partial autocorrelations using the recursive relation from [30]. The adjacent panel on the tab will visualize the mean plots with their accompanying standard errors for both treatment arms. For

dichotomous outcomes, users can specify the threshold value such that values greater than the threshold will be 1 and 0 otherwise, corresponding to a "clinical failure" or a "clinical success". Figure 3.1 shows the mean settings tab for CITIES.

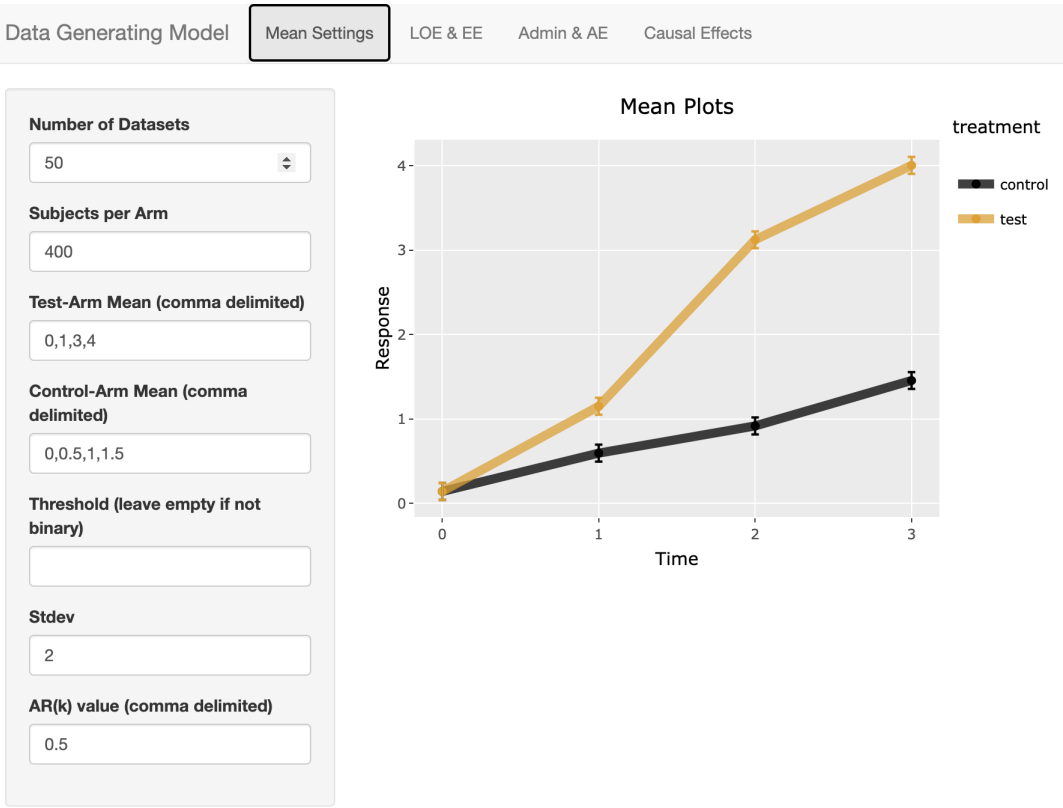


Figure 3.1. Mean Settings Tab for CITIES

3.2.2 LOE & EE: Lack Of Efficacy & Excess Efficacy Curves

Probability of discontinuing due to LOE & EE are defined as piecewise linear functions that plateau beyond some specified thresholds. Define Pr(DC | LOE) to be the probability of discontinuing due to LOE, $p_{\text{loe,max}}$ to be the ceiling value of discontinuing due to LOE, d to be the difference between the potential outcome and the corresponding baseline measurement, $y_{\text{L,loe}}$ to be the lower threshold and $y_{\text{U,loe}}$ to be the upper threshold. Then

$$\Pr(\text{DC} \mid \text{LOE}) = f_{\text{loe}}(d) = \begin{cases} p_{\text{loe,max}} & d \leq y_{\text{L,loe}} \\ m_{\text{loe}} d + b_{\text{loe}} & y_{\text{L,loe}} < d \leq y_{\text{U,loe}} \\ 0 & d > y_{\text{U,loe}} \end{cases} \tag{3.2}$$

where $p_{\text{loe,max}} \in [0,1], m_{\text{loe}} = \frac{0-p_{\text{loe,max}}}{y_{\text{U, loe}}-y_{\text{L,loe}}}$ and $b_{\text{loe}} = 0 - m_{\text{loe}} y_{\text{U,loe}}$. Similarly, define Pr(DC | EE) to be the probability of discontinuing due to EE, $p_{\text{ee,max}}$ to be the ceiling value of discontinuing due to EE, d to be the difference between the potential outcome and the corresponding baseline measurement, $y_{\text{L,ee}}$ to be the lower threshold and $y_{\text{U,ee}}$ to be the upper threshold. Then

$$\Pr(\text{DC} \mid \text{EE}) = f_{\text{ee}}(d) = \begin{cases} 0 & d \leq y_{\text{L,ee}} \\ m_{\text{ee}} d + b_{\text{ee}} & y_{\text{L,ee}} < d \leq y_{\text{U,ee}} \\ p_{\text{ee,max}} & d > y_{\text{U,ee}} \end{cases} \tag{3.3}$$

where $p_{\text{ee, max}} \in [0,1], m_{\text{ee}} = \frac{p_{\text{ee, max}}-0}{y_{\text{U, ee}}-y_{\text{L, ee}}}$ and $b_{\text{ee}} = 0 - m_{\text{ee}} y_{\text{L, ee}}$. On the LOE & EE tab, users will first specify if higher values are better or not. In diabetes, lower blood glucose is a desirable outcome and is often measured by reductions in glycated hemoglobin or HbA1c [99]. Conversely, in Alzheimer's Disease trials, there are cognition rating scales used to measure the patient's disease status, and higher cognition scores are of interest [100]. Contingent on that, the LOE and EE will reflect this dynamically, since Pr(DC | LOE) will be a decreasing function while and Pr(DC | EE) will be an increasing function when higher values are better and vice-versa. Figure 3.2 shows the LOE & EE tab for CITIES.

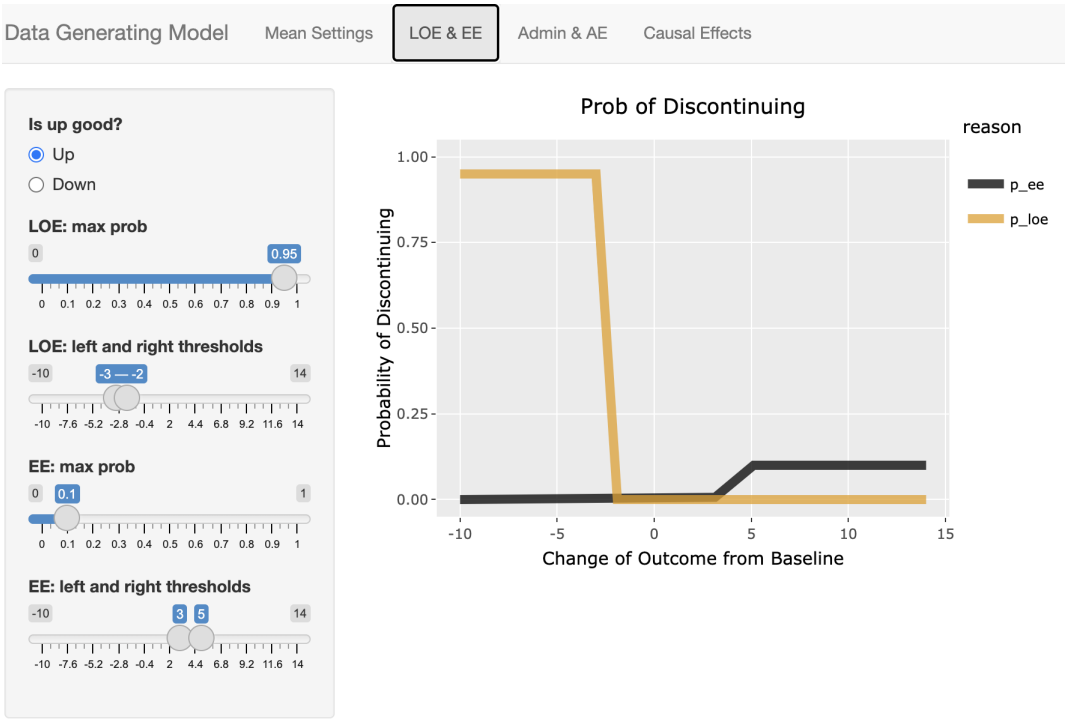


Figure 3.2. LOE & EE Tab for CITIES simulator

3.2.3 AE & Admin: Adverse Events & Administrative Curves

The discontinuation due to AE behavior is modelled as an increasing linear function which starts at the origin and extends to the Cartesian point of the maximum probability of an AE occurring on treatment arm t at the final time point J: $(J, p_{\text{ae, max},t})$. The probability of experiencing an AE on each study treatment is often summarized in some way when reporting the results of a clinical trial.

Not all patients who experience an AE discontinue their randomized study treatment. Thus, the model should incorporate not only the probability of experiencing an adverse event, but the conditional probability of discontinuing study treatment given that the patient experiences an AE. Since patient behavior and responses may differ across the treatment arms, the simulator allows for different different discontinuation probability given an AE for test and control. The joint probability of discontinuing and occurrence of an AE on the treatment arm t is the product of the probability of discontinuing due to an AE on arm t and the probability of an occurrence of an AE on the corresponding treatment arm, i.e. $\Pr(\text{DC, AE} \mid \text{t}) = \Pr(\text{DC} \mid \text{AE, t}) \times \Pr(\text{AE} \mid \text{t})$.

Patients may also discontinue due to administrative reasons. These are reason not related to study treatment such as (a) personal events that preclude further participation in a clinical trial (e.g. moving to a new location, pregnancy, change in marital status) or (b) patients voluntary unwillingness to continue to participate in frequent doctor visits or other burdens related to onerous clinical evaluations of the patient. As such the probability of discontinuation generally increases with the duration of the trial – at the beginning of the study or for shorter duration trials, few such administrative discontinuations occurs, but with studies lasting years in some cases, such discontinuations become more frequent with time. The probability of discontinuing due to administrative reasons is described using a linear function of time from the origin to the maximum probability of discontinuing due to administrative reasons at the final time point.

On the Admin & AE tab, users will specify the maximum probabilities of an AE occurring as well as the probability of discontinuing if an adverse event occur on both arms. In addition, users will also have to specify the maximum probability of discontinuing due to administrative reasons. Figure 3.3 shows the AE tab for CITIES.

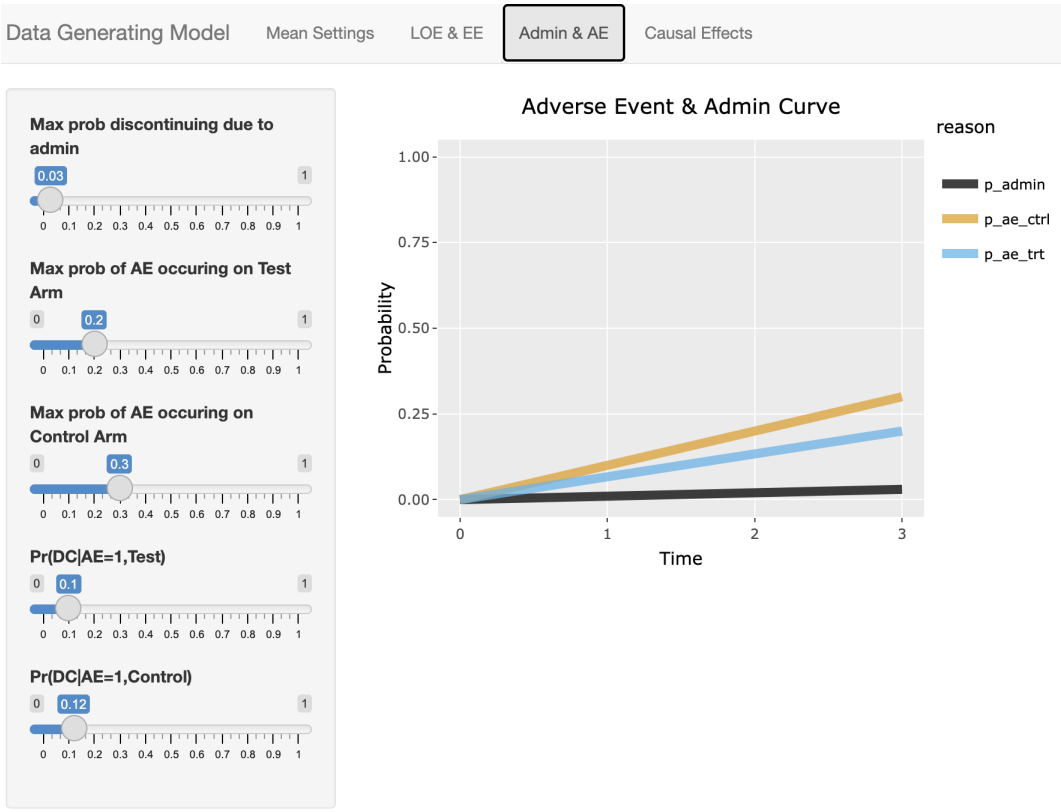


Figure 3.3. AE Tab for CITIES

Incorporating all individual components, CITIES proceeds by first generating all potential outcomes y_{ijt} based on user specifications on both the test and control arms, conditional on the baseline measurement. Subsequently, probability of discontinuations are generated due to ICEs (AE, LOE, EE, Admin). Discontinuations are then induced in the generated potential outcomes, yielding the true data. Finally, each subject is randomized to either the test or control arm to get realized outcomes or what one would actually observe in a clinical trial.

3.2.4 Causal Effect

We follow the definition of an estimand on the treatment effect from *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials* [86], which is "estimated by comparing the outcomes in a group of subjects on the treatment to those in a similar group of subjects on the control". The natural causal estimand here would be comparing outcomes across treatment arms for those who are able to adhere to their treatment assignment, i.e. adhere average causal effect (AdACE). Following the notation of Qu, Luo, and Ruberg [87], define $A(t) = I\{\text{adhered to treatment } t\}$, where $I\{.\}$ is an indicator function. Then

$S_{++} = \{A(0) = A(1) = 1\}$ or 'Adhere_both' is the stratum comprising those who were able to adhere completely to their treatment assignments on both treatment arms and S or 'No_dc' is the hypothetical stratum where all subjects adhered to their treatment assignments and never discontinued on both arms. With Y being the potential outcome, the corresponding AdACE for these two strata at time j are defined as follows:

$$ACE(S_{++})_{\mathrm{j}} = \mathbb{E}_{\mathrm{j}}[Y|t = 1, S_{++}] - \mathbb{E}_{\mathrm{j}}[Y|t = 0, S_{++}] \tag{3.4}$$

$$ACE(S)\mathrm{j} = \mathbb{E}_{\mathrm{j}}[Y|t = 1, S] - \mathbb{E}_{\mathrm{j}}[Y|t = 0, S] \tag{3.5}$$

On the causal effects tab, the ACEs are visualized at each time point for the two different strata with their corresponding standard errors averaged over the data replicates. Users can also bookmark the simulation input parameters to be used in a later time. Once the Submit action button is clicked, a progress bar on the bottom right of the window will show at what iteration is the simulator at. Figure 3.4 shows the ACEs on the causal effects tab for CITIES.

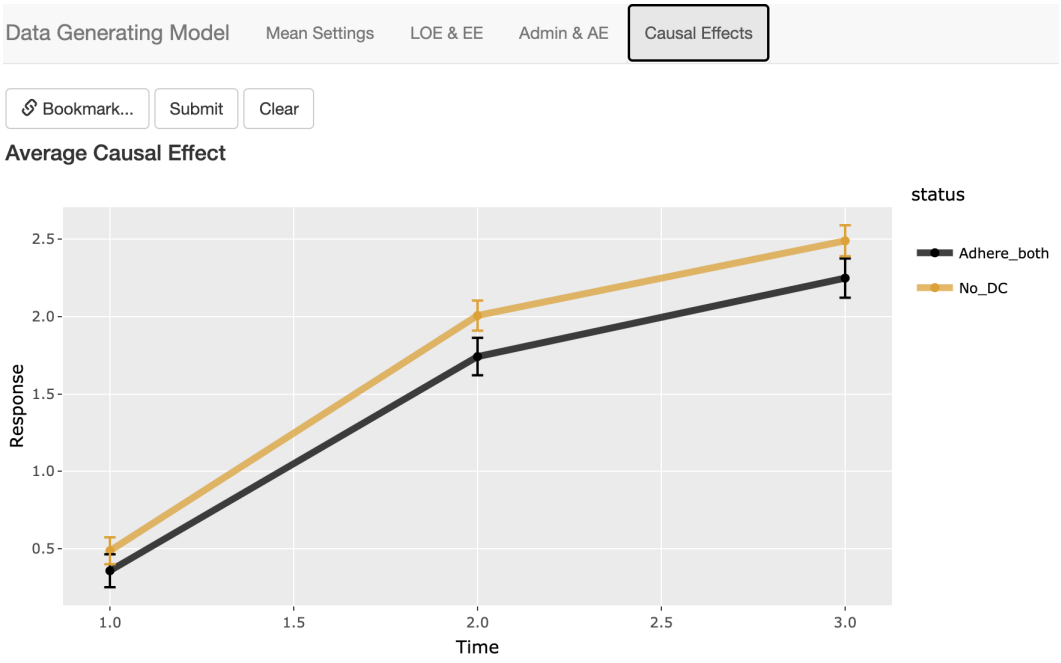


Figure 3.4. Average causal effect for CITIES

The final visual in CITIES summarizes the percentage discontinuation at each time point and their corresponding reasons, reflected in figure 3.5. This is especially useful when trying to mimic real clinical trials without having access to the real data.

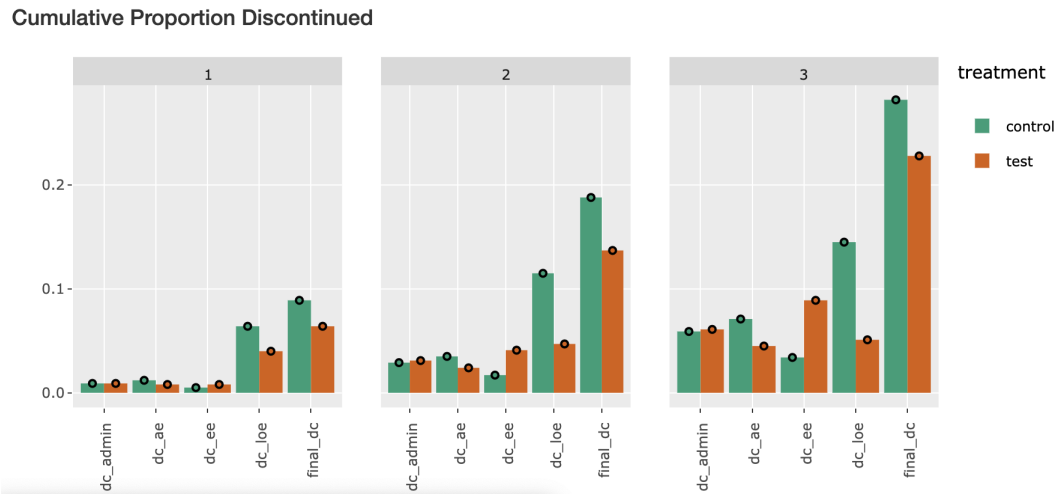


Figure 3.5. Percentage missing of data for CITIES

3.3 Real Data Example

3.3.1 Canagliflozin

We will demonstrate CITIES on a canagliflozin monotherapy clinical trial [101]. Canagliflozin is a sodium glucose co-transporter 2-inhibitor that was developed for type 2 diabetes mellitus (T2DM). In this 26-week randomized, double-blinded, placebo-controlled phase 3 trial, each of the 584 subjects received either canagliflozin 100 mg, 300 mg or placebo once daily. In this simulation, we compared the 100mg dose and placebo. The primary endpoint was the change from baseline in haemoglobin A1c (HbA1c) at week 26. Input parameters for mean settings tab in CITIES in Figure 3.6 were chosen based on the raw means and standard errors of the response from Graph B in Figure 2 of the study. Both the LOE & EE (Figure 3.7) and Admin & AE (Figure 3.8) tabs were populated based on the study flow diagram or Figure 1 from Stenlöf, Cefalu, Kim, *et al.* [101]. Further, we checked the percentage discontinuation from our simulated data (Figure 3.10) against Figure 1 from the same study to validate how reasonable our input parameter settings were.

The estimated primary endpoint ACE using CITIES for the canagliflozin 100mg relative to the placebo is -0.703 for the S_{++} stratum (Table 3.1). This is slightly less in magnitude then the estimated treatment effect from the study of -0.77 in Graph A from Figure 2 , where analysis was focused on the modified intent-to-treat (mITT) population comprising all randomized subjects who had received at least 1 dose of the study and The Last Observation Carried Forward (LOCF) procedure was used to impute missing data. The trends of the secondary endpoints in Figure

3.9 runs in tandem with the mITT population from the original study. The estimated treatment effect is quite close to the primary endpoint ACE for the S stratum that assumed no participant discontinued. We can see that the CITIES mimics the real data easily and can be used to simulate clinical trials to test estimands and estimators more effectively since we know the true AdACE in the presence of ICEs.

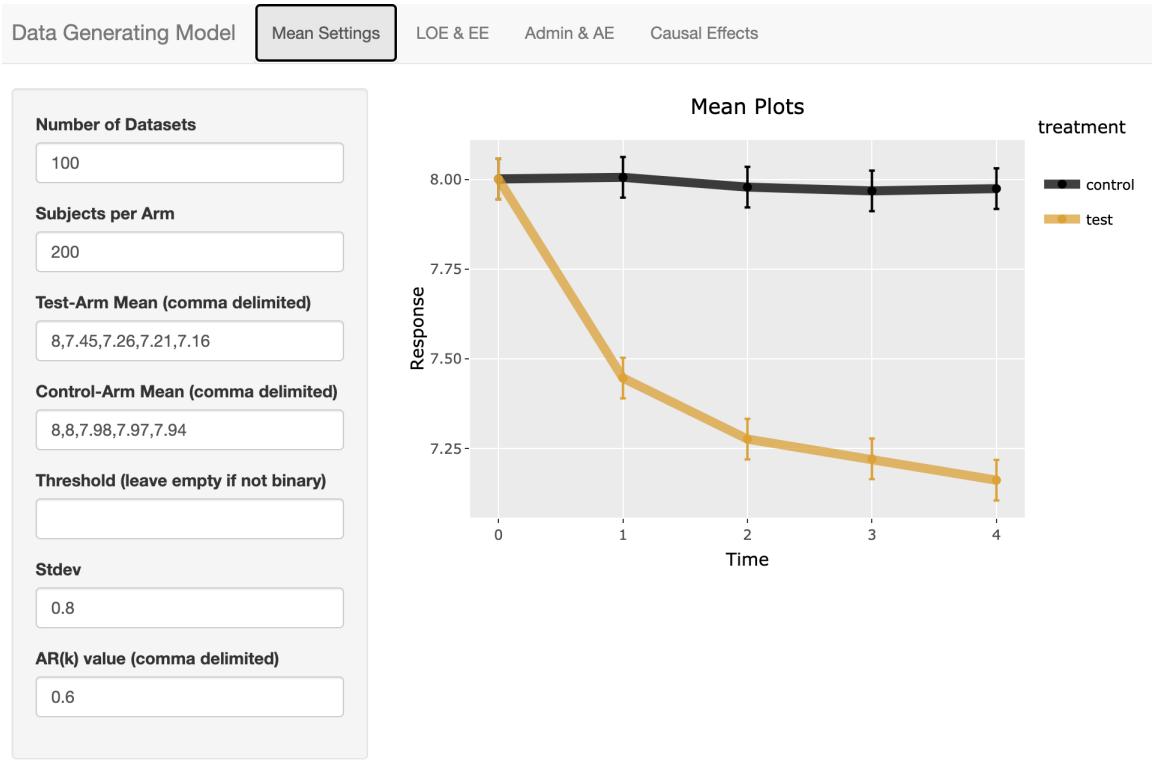


Figure 3.6. CITIES mean settings tab for canagliflozin

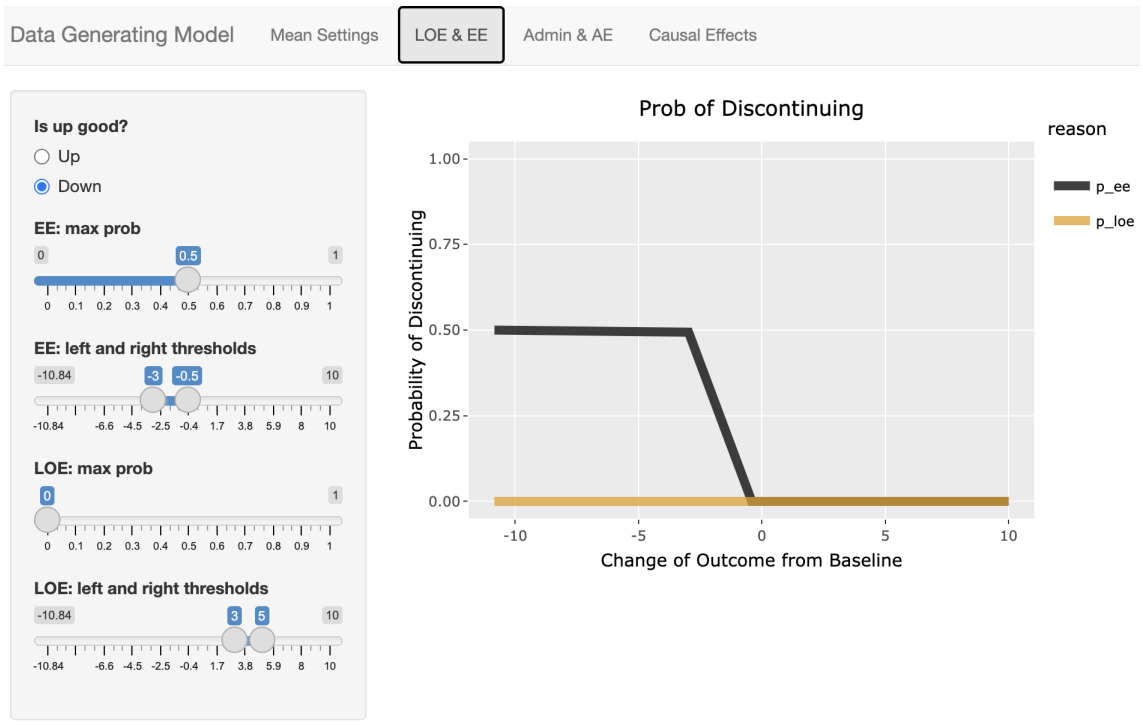


Figure 3.7. CITIES LOE & EE tab for canagliflozin

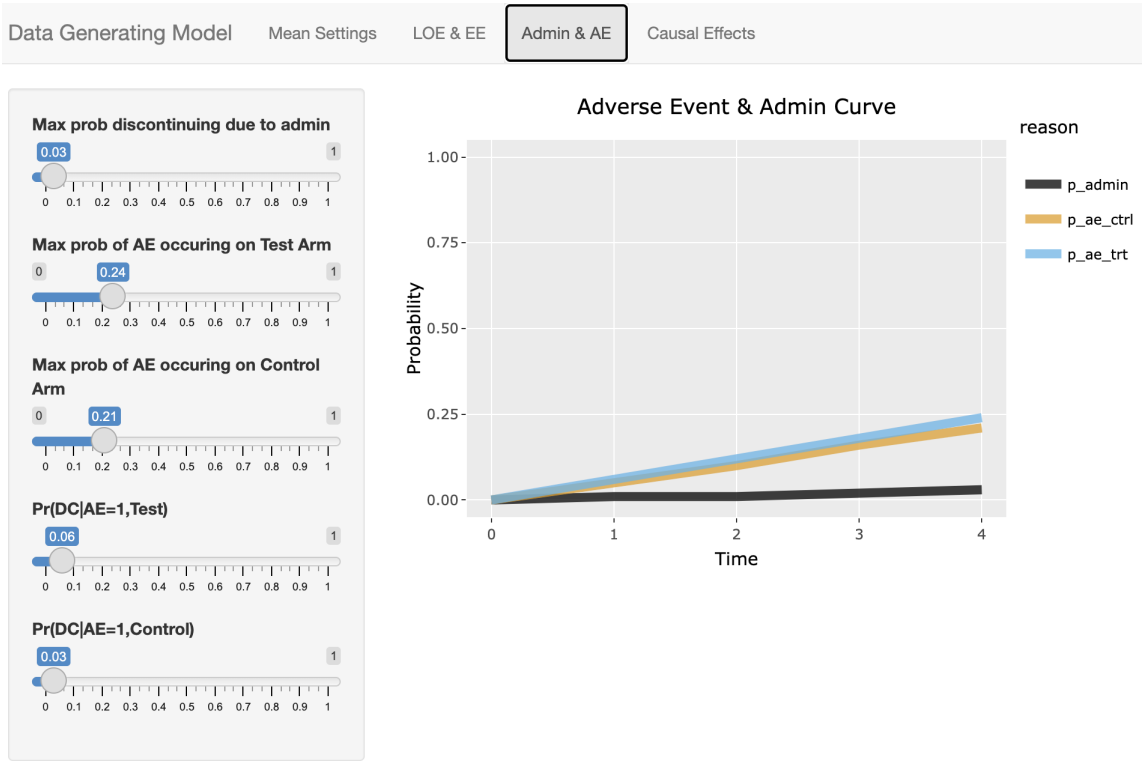


Figure 3.8. CITIES Admin & AE tab for canagliflozin

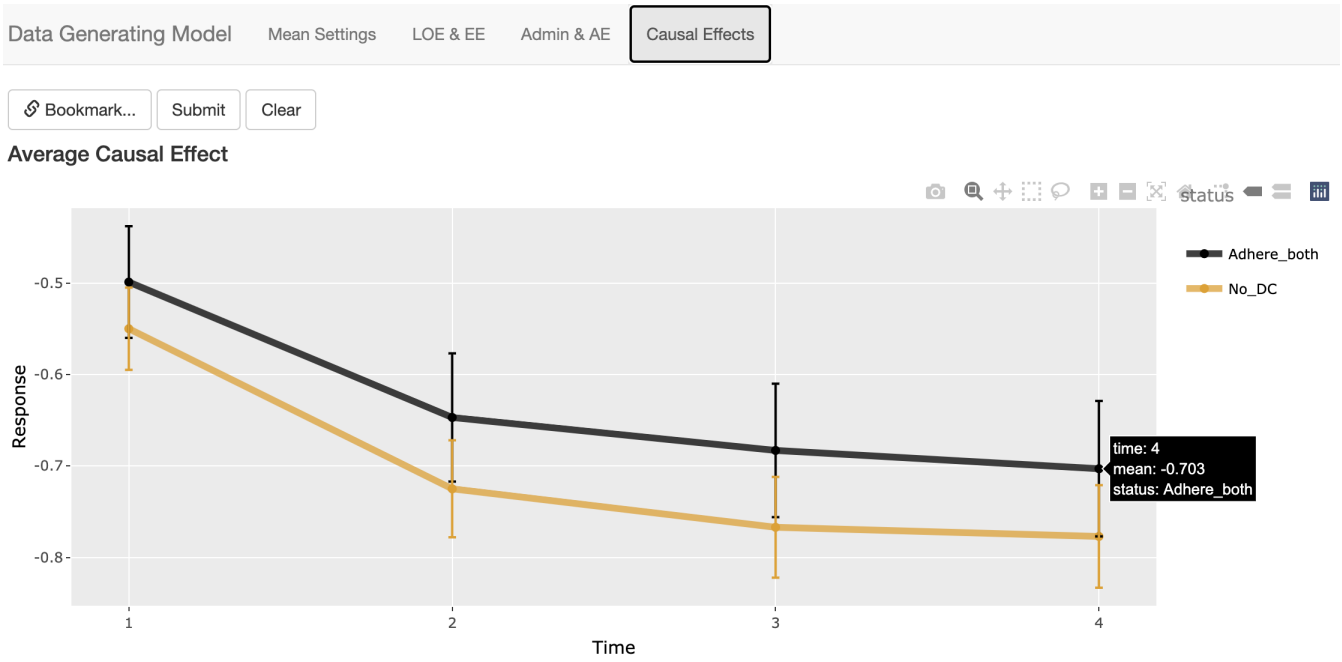


Figure 3.9. Average causal effect for canagliflozin

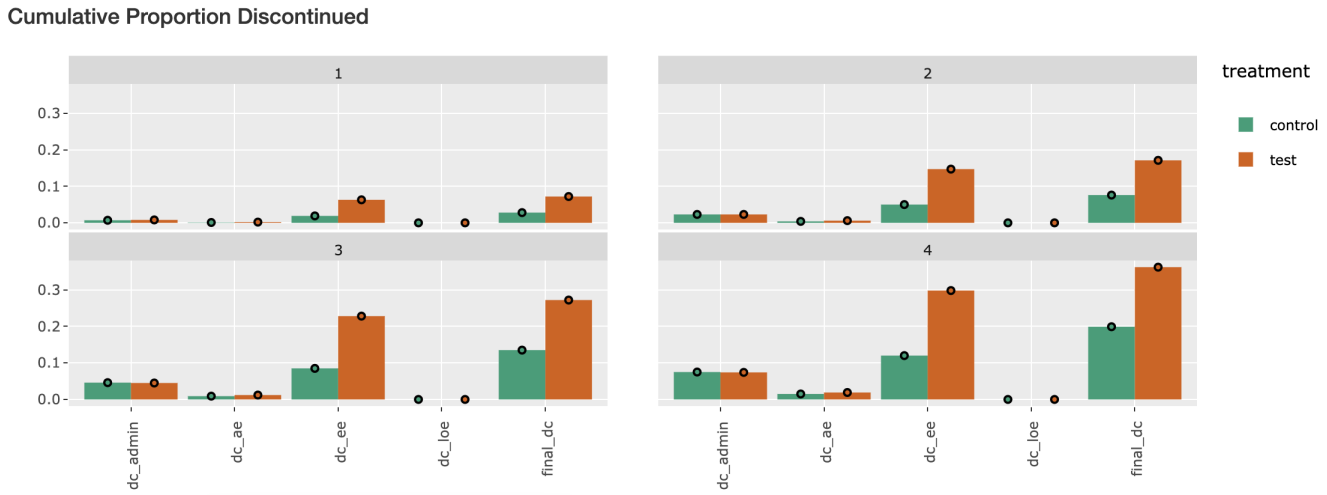


Figure 3.10. Percentage missing simulated data for canagliflozin study

Table 3.1. Average Causal Effect for canagliflozin for each time point using CITIES. The highlighted rows are the ACEs for the primary endpoints.

Stratum	Time	ACE
S_{++}	4	-0.703 (0.076)
S	4	-0.777 (0.056)
S_{++}	3	-0.685 (0.074)
S	3	-0.767 (0.055)
S_{++}	2	-0.647 (0.071)
S	2	-0.725 (0.053)
S_{++}	1	-0.501 (0.062)
S	1	-0.55 (0.045)

3.3.2 Donanemab

Another application of CITIES is on the donanemab trial in early Alzheimer's diseases [102]. Donanemab is an antibody that targets a modified form of Aβ deposits, a peptide that has been suggested to grow in correlation to the progression of Alzheimer's disease. In this phase 2 trial, 257 patients were enrolled with 131 assigned to receive the active treatment and the remaining 126 were assigned to the placebo group. Donanemab was administered intravenously according to a carefully designed titration scheme: 700 mg for the first three doses and 1400 mg beyond that. All patients had early symptomatic Alzheimer's disease who had amyloid and tau deposits on their Positron-Emission Tomography (PET). The primary endpoint was the change in Integrated Alzheimer's Disease Rating Scale (iADRS; score that ranges from 0 to 144 with lower scores indicating greater functional and cognitive impairment) from baseline. The mean settings tab for CITIES (Figure 3.11) was populated based on Graph A in Figure and that the baseline measurement for all patients on both arms was 106. The LOE & EE and Admin & AE tabs in Figures 3.12 and 3.13 respectively were populated based on Table 2 from the original publication. As before, we checked the percentage discontinuation from our simulated data (Figure 3.15) against Table 2 from the same study to see how well our proportions of missingness match up with the original study.

The estimated primary endpoint ACE using CITIES for donanemab relative to the placebo is 3.859 for the S_{++} stratum (Table 3.2). This is larger in magnitude then the estimated treatment effect of 3.2 from the study in Graph A from Figure 2, where analysis was focused on the mITT

population comprising all randomized subjects who had received at least 1 dose of the study and a Mixed Model for Repeated Measures (MMRM) was used to estimate the Least-Square Means, while controlling for fixed effects listed in the study. Time was treated as a discrete variable and incorporated into the random effects of the MMRM using an unstructured covariance structure. Here we see that CITIES provides an interactive and transparent platform where users can simulate clinical trials to test estimands and estimators based on real clinical trials.

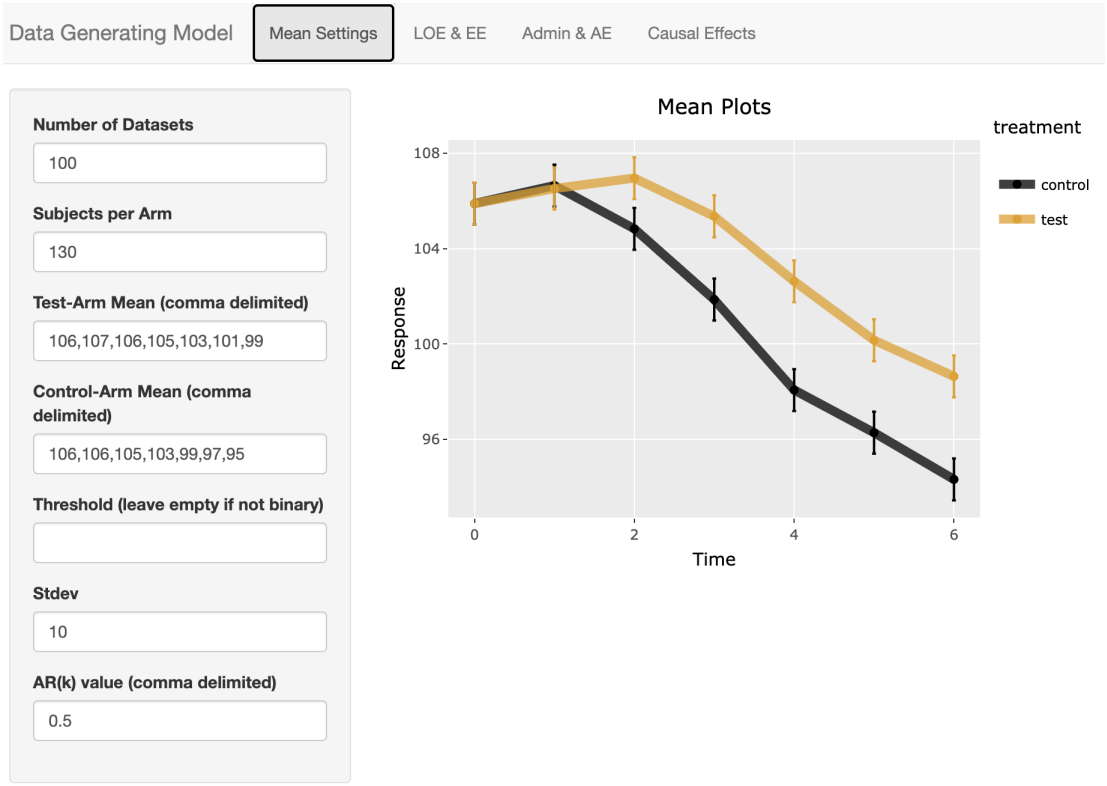


Figure 3.11. CITIES mean settings tab for donanomeb

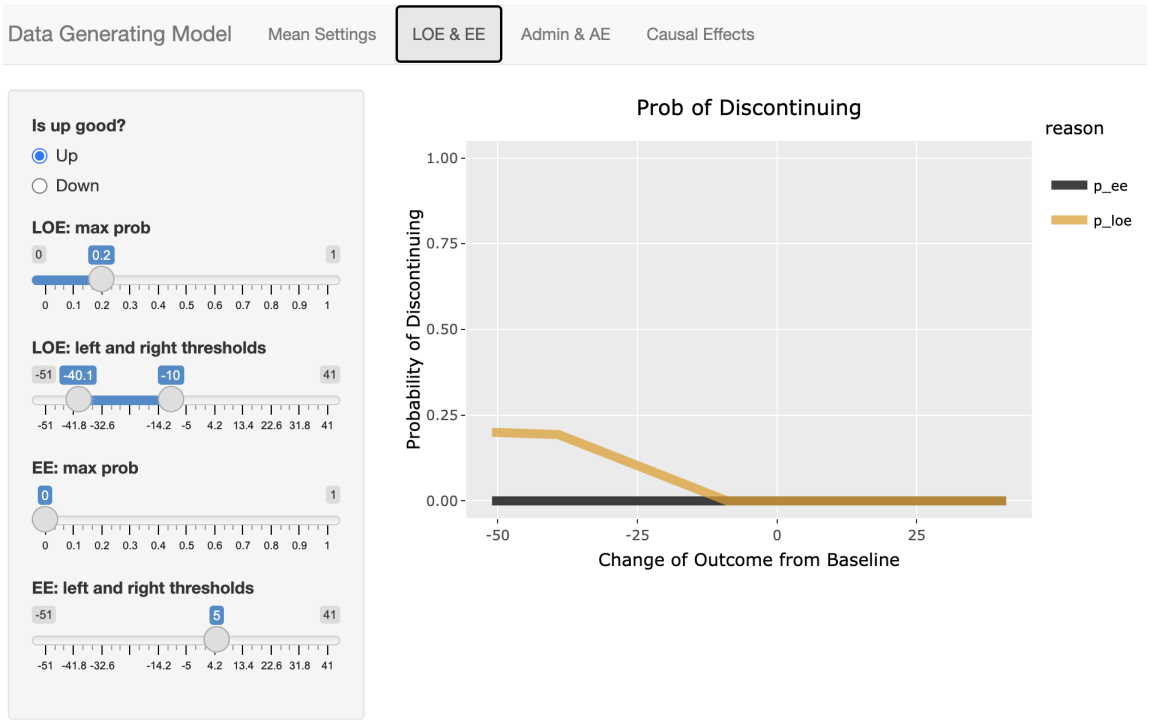


Figure 3.12. CITIES LOE & EE tab for donanomeb

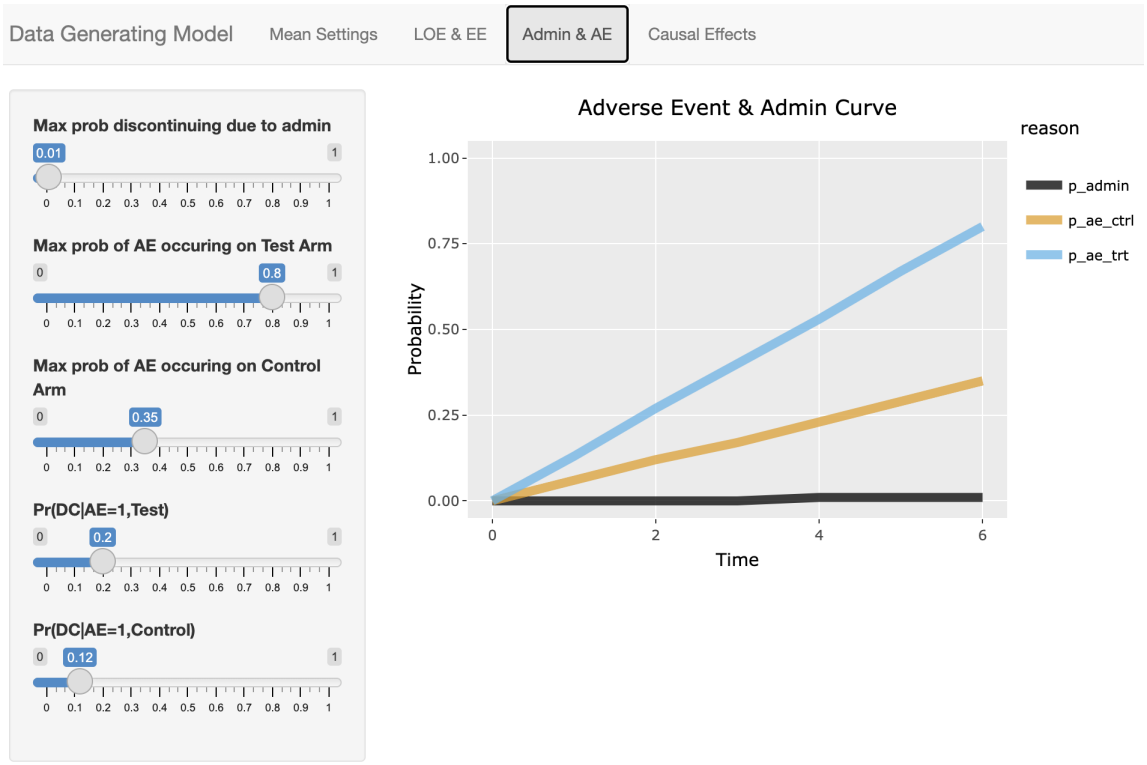


Figure 3.13. CITIES Admin & AE tab for donanomeb

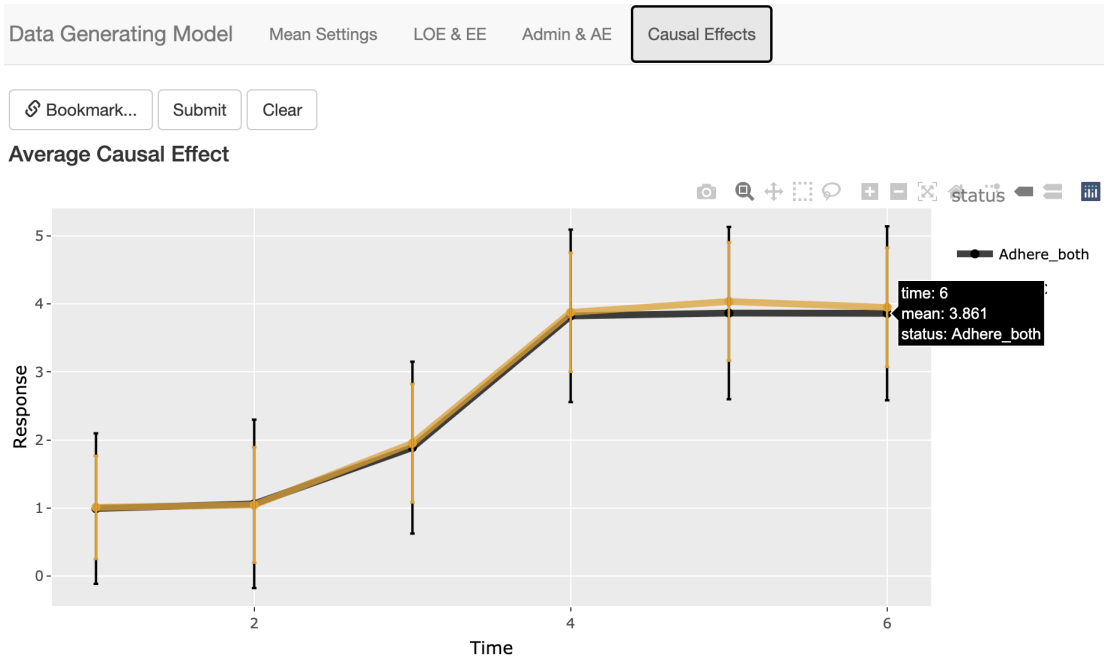


Figure 3.14. Average causal effect for donanomeb

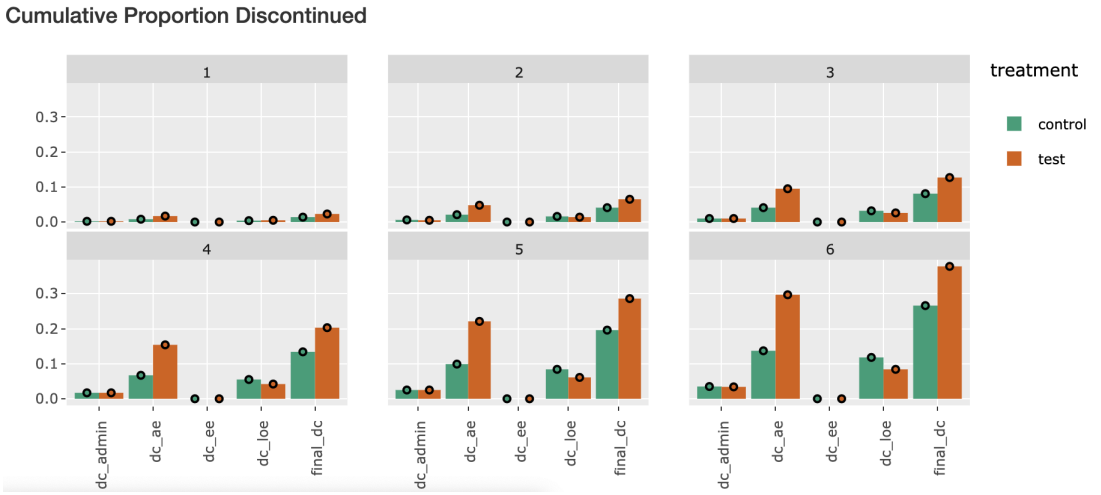


Figure 3.15. Percentage missing simulated data for donanomeb study

Table 3.2. Average Causal Effect for donanomeb for each time point using CITIES. The highlighted rows are the ACEs for the primary endpoints.

Stratum	Time	ACE
S_{++}	6	3.859 (1.277)
S	6	3.949 (0.876)
S_{++}	5	3.86 (1.265)
S	5	4.035 (0.87)
S_{++}	4	3.82 (1.268)
S	4	3.875 (0.875)
S_{++}	3	1.886 (1.263)
S	3	1.952 (0.87)
S_{++}	2	1.046 (0.848)
S	2	1.065 (1.236)
S_{++}	1	0.991 (1.106)
S	1	1.01 (0.756)

3.4 Discussion

We have created CITIES that has three major advantages over any existing clinical trial simulation engines: (1) it generates efficacy data in the potential outcomes framework consistent with the ICH E9(R1) definition of a treatment effect; (2) it also generates potential outcomes for adverse events; and (3) it incorporates a set of realistic study treatment discontinuation models for the primary reasons for treatment discontinuation seen in real clinical trials. It is the combination of the three unique elements that we believe can better serve to generate simulated clinical trial data sets for the purpose of developing, implementing and comparing different statistical estimation methods for different estimands of interest. We are continuing to work on this model and refine it with the intent to capture additional complexity of clinical trials while maintaining a parsimonious model with a modest number of parameter inputs for the user.

A possible extension to the current simulator would be to include covariates in the means of the distributions and in inducing discontinuation in the data. For example, older patients may have a higher probability of discontinuing, or patients with worse baseline severity might have a lower probability of discontinuing (i.e., more severe patients might be more keen to get relief and

therefore stay in the trial longer). This facet was not included in the current simulator since unlike ICEs, there are infinite ways in which baseline covariates affect patient behaviour as well as the number of such baseline covariates could be involved (e.g., is there any influence of being both old and very sick?). Such added complexity to the model must be balanced with keeping CITIES easy to use.

The CITIES simulator is a convenient web application that can generate data using potential outcomes in the tripartite framework and provide causal estimates while providing an intuitive mechanism to generate discontinuation using ICEs. As is true with all fields of science, communicating ideas and sharing information embodied in a model can be challenging. The CITIES allows users to share and present information dynamically with ease and transparency.

4. CONCLUDING REMARKS

Current clinical trials involve novel measures of disease progression that are complicated and often not continuous in nature, such as the ones we have considered in this study where the outcomes are binary. The task of modelling these outcomes are further burdened with intricate serial correlations that can take form with a growing wealth of repeated measurements. Formulating the correlation structure when analyzing the data requires careful thought in what we can parameterize while maintaining an acceptable degree of model parsimony. In this study, the structured binary repeat measurements compel us to use a general autoregressive process that capitalizes on the discrete and systematic nature of the study while affording flexibility.

We developed two methodologies for addressing these challenges. The BGLAM allows users to incorporate nested autoregressive processes using non-informative priors while maintaining marginal interpretations of the coefficients. Through extensive simulations, we have demonstrated that BGLAM is more capable in testing for treatment effects in clinical trials with repeated binary measurements and systemic autoregressive processes. Although the coefficients in BGLAM have favorable marginal interpretations, the model lacks the feature to incorporate heterogeneity across time and in the autoregressive process and is computationally consuming. To overcome this, we introduced the PGLAM model that is both more computationally efficient and can integrate additional layers of variability via random effects and heterogeneity across time. We showed through simulations that our model better recovers these layers of variability which translates to better power and coverage when testing or treatment effects for studies with small to moderate sample sizes and large number of repeat measurements. A natural implementation of BGLAM and PGLAM are in causal models with binary repeat measurements such as Principal Stratification where latent structures are imputed at each iteration which would warrant a correlation structure that may be different across the different latent strata. With smaller number of repeat measurements, users can revert to the original BMLR implementation in [4] and have an unstructured corrrelation structure where each correlation input uniformly varies in the [-1,1] space.

An added layer of complication to clinical trials with repeat measurements or multiple timepoints is treatment nonadherence. Inevitably, when some patients discontinue the treatment intervention due to intercurrent events such as administrative reasons, excess efficacy, lack of efficacy or adverse events, these discontinuations which occur as an implicit function of unobserved outcomes will potentially mar inferred treatment effects from models that do not account for these

discontinuations. As such, the pharmaceutical industry has invested and developed a myriad of models that address these discontinuations due to intercurrent events. However, it is consequently very difficult to compare model performances of these competing models as real-life clinical trial data cannot be so easily shared publicly.

To address these challenges, we developed the DGM, an R shiny app, that allows users to dictate patient compliance via varying sources of discontinuity with different functional behaviors and generate data to get a more holistic understanding of the operating characteristics of the investigational treatment obtained by different models for requested estimands with multiple endpoints. The DGM is a convenient app that updates interactively and provides results synchronously, allowing users to share their results and data input settings to generate the data without actually having to share their clinical trials data. This convenience is met with some limitations. For instance, although covariate information can be readily integrated into the DGM, this would undermine the simplicity and directness of our simulator since there would be infinite ways for different covariates to be integrated into the simulation process. Users are more than welcome to use the code that is readily available upon request and amend as they see fit to meet their respective needs.

REFERENCES

[1] M. B. M. B. K. Gawarammana and M. R. Sooriyarachchi, "Comparison of methods for analyzing binary repeated measures data: A simulation-based study (comparison of methods for binary repeated measures)," eng, *Communications in Statistics - Simulation and Computation*, vol. 46, no. 3, pp. 2103–2120, 2017, ISSN: 0361-0918. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03610918.2015.1035445.

[2] W. W. Stroup, *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*, eng, ser. Texts in Statistical Science. CRC Press, Taylor & Francis Group, 2013, ISBN: 1439815127.

[3] P. Mccullagh, "Regression models for ordinal data," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 42, no. 2, pp. 109–127, 1980, ISSN: 0035-9246.

[4] S. M. O'Brien and D. B. Dunson, "Bayesian multivariate logistic regression," *Biometrics*, vol. 60, no. 3, pp. 739–746, 2004, ISSN: 0006-341X.

[5] K. Liang and S. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986, ISSN: 00063444.

[6] R. L. Prentice, "Correlated binary regression with covariates specific to each binary observation," eng, *Biometrics*, vol. 44, no. 4, pp. 1033–1048, 1988, ISSN: 0006341X.

[7] S. R. Lipsitz, N. M. Laird, and D. P. Harrington, "Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association," eng, *Biometrika*, vol. 78, no. 1, pp. 153–160, 1991, ISSN: 00063444.

[8] V. Carey, S. L. Zeger, and P. Diggle, "Modelling multivariate binary data with alternating logistic regressions," eng, *Biometrika*, vol. 80, no. 3, pp. 517–526, 1993, ISSN: 00063444.

[9] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society: Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972. DOI: 10.2307/2344614. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.2307/2344614. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2344614.

[10] A. Hubbard E., J. Ahern L., N. Fleischer Van Der, M. Laan A., S. Lippman A., N. Jewell A., T. Bruckner A., and W. Satariano A., "To gee or not to gee: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health," *Epidemiology*, vol. 21, no. 4, pp. 467–474, 2010, ISSN: 1044-3983.

[11] C. Mcculloch, "Maximum likelihood algorithms for generalized linear mixed models," *Journal of the American Statistical Association*, vol. 92, no. 437, pp. 162–170, 1997, ISSN: 0162-1459.

[12] J. C. Pinheiro and D. M. Bates, "Approximations to the log-likelihood function in the nonlinear mixed-effects model," *Journal of Computational and Graphical Statistics*, vol. 4, no. 1, pp. 12–35, 1995, ISSN: 1061-8600.

[13] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, eng. United States Department of Commerce, 1972.

[14] SAS Institute Inc, *SAS/STAT 9.2 User's Guide*, English, 1st edition. Cary, NC: SAS Publishing, 2008, ISBN: 1555443761.

[15] L. Ferrucci, *The baltimore longitudinal study of aging (blsa): A 50-year-long journey and plans for the future*, eng, 2008.

[16] Ohio State University Center for Human Resource Research, *National longitudinal surveys of labor market experience, 1966-1992*, 2008.

[17] C. Ekinsmyth, "Large-scale longitudinal studies: Their utility for geographic enquiry," English, *Area*, vol. 28, no. 3, pp. 358–372, 1996, ISSN: 0004-0894.

[18] J. J. Jansen, H. C. J. Hoefsloot, H. F. M. Boelens, J. van der Greef, and A. K. Smilde, "Analysis of longitudinal metabolomics data," *Bioinformatics*, vol. 20, no. 15, pp. 2438–2446, 2004, ISSN: 1367-4803.

[19] N. Nooraee, F. Abegaz, J. Ormel, E. Wit, and E. R. Van Den Heuvel, "An approximate marginal logistic distribution for the analysis of longitudinal ordinal data," *Biometrics*, vol. 72, no. 1, pp. 253–261, 2016, ISSN: 0006-341X.

[20] R. Hirk, K. Hornik, L. Vana, and A. Genz, *Mvord: Multivariate ordinal regression models*, 2019.

[21] E. Paul, A. K. Maity, and R. Maiti, "Bayesian comparative study on binary time series," eng, *Journal of Statistical Computation and Simulation*, vol. 88, no. 14, pp. 2811–2826, 2018, ISSN: 0094-9655. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00949655.2018.1488256.

[22] G. Casella, *Statistical Inference*, eng, 2nd, ser. Duxbury Advanced Series. Australia ; Pacific Grove, CA: Duxbury/Thomson Learning, 2002, ISBN: 0534243126.

[23] P. S. Laplace, "Memoir on the probability of the causes of events," eng, *Statist. Sci.*, vol. 1, no. 3, pp. 364–378, 1986, ISSN: 0883-4237.

[24] J. Raphson, *Analysis quationum universalis, seu, Ad quationes algebraicas resolvendas methodus generalis*, lat, second edition, ser. Early English books online. Londini: Typis T. Braddyll, prostant venales apud Johannem Taylor ..., 1697. [Online]. Available:

.

[25] J. H. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 669–679, 1993, ISSN: 0162-1459.

[26] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970, ISSN: 0006-3444. DOI: 10.1093/biomet/57.1.97. eprint: http://oup.prod.sis.lan/biomet/article-pdf/57/1/97/23940249/57-1-97.pdf. [Online]. Available: https://doi.org/10.1093/biomet/57.1.97.

[27] O. Barndorff-Nielsen and G. Schou, "On the parametrization of autoregressive models by partial autocorrelations," *Journal of Multivariate Analysis*, vol. 3, no. 4, pp. 408–419, 1973, ISSN: 0047-259X. DOI: https://doi.org/10.1016/0047-259X(73)90030-4. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0047259X73900304.

[28] F. L. Ramsey, "Characterization of the partial autocorrelation function," eng, *The Annals of Statistics*, vol. 2, no. 6, pp. 1296–1301, 1974, ISSN: 00905364.

[29] T. W. (W. Anderson, *An Introduction to Multivariate Statistical Analysis*, eng, 3rd, ser. Wiley Series in Probability and Statistics. Hoboken, N.J.: Wiley-Interscience, 2003, ISBN: 0471360910.

[30] H. Joe, "Generating random correlation matrices based on partial correlations," eng, *Journal of Multivariate Analysis*, vol. 97, no. 10, pp. 2177–2189, 2006, ISSN: 0047-259X.

[31] Y. Wang and M. Daniels, "Bayesian modeling of the dependence in longitudinal data via partial autocorrelations and marginal variances," eng, *Journal of Multivariate Analysis*, vol. 116, 2013, ISSN: 0047-259X.

[32] M. Daniels and M. Pourahmadi, "Modeling covariance matrices via partial autocorrelations," eng, *Journal of Multivariate Analysis*, vol. 100, no. 10, pp. 2352–2363, 2009, ISSN: 0047-259X.

[33] A. Gelman, *Bayesian data analysis*, eng, Third edition.., ser. Chapman & Hall/CRC texts in statistical science. CRC Press, 2014, ISBN: 1439840954.

[34] G. T. Wilson, "On the use of marginal likelihood in time series model estimation," eng, *Journal of the Royal Statistical Society. Series B, Methodological*, vol. 51, no. 1, pp. 15–27, 1989, ISSN: 0035-9246.

[35] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," eng, *Journal of the Royal Statistical Society. Series B, Statistical methodology*, Journal of the Royal Statistical Society Series B, vol. 64, no. 4, pp. 583–639, 2002, ISSN: 1369-7412.

[36] K. E. Lee, N. Sha, E. R. Dougherty, M. Vannucci, and B. K. Mallick, "Gene selection: A bayesian variable selection approach," eng, *Bioinformatics (Oxford, England)*, vol. 19, no. 1, pp. 90–97, 2003, ISSN: 1367-4803.

[37] A. Linde, "Dic in variable selection," eng, *Statistica Neerlandica*, Statistica Neerlandica, vol. 59, no. 1, pp. 45–56, 2005, ISSN: 0039-0402.

[38] A. Berg, R. Meyer, and J. Yu, "Deviance information criterion for comparing stochastic volatility models," eng, *Journal of business & economic statistics*, vol. 22, no. 1, pp. 107–120, 2004, ISSN: 0735-0015.

[39] V. J. Carey, T. Lumley, and B. Ripley., *Gee: Generalized estimation equation solver*, R package version 4.13-20, 2019. [Online]. Available: https://CRAN.R-project.org/package=gee.

[40] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," eng, *Annual review of clinical psychology*, vol. 4, no. 1, pp. 1–32, 2008, ISSN: 1548-5943.

[41] C. Vincent, E. Auger, V. Lavoie, M. Besemann, N. Champagne, G. Belleville, E. Beland, E. Bernier-Banville, and J. Bourassa, "Service dog schools for ptsd as a tertiary prevention modality: Assessment based on assistance dogs," *Edelweiss: Psychiatry Open Access*, pp. 29–41, Jun. 2019. DOI: 10.33805/2641-8991.119.

[42] T. K. Crowe, V. Sanchez, A. Howard, B. Western, and S. Barger, "Veterans transitioning from isolation to integration: A look at veteran/service dog partnerships," eng, *Disability and rehabilitation*, vol. 40, no. 24, pp. 2953–2961, 2018, ISSN: 0963-8288.

[43] J. A. Cranford, P. E. Shrout, M. Iida, E. Rafaeli, T. Yip, and N. Bolger, "A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably?" eng, *Personality & social psychology bulletin*, vol. 32, no. 7, pp. 917–929, 2006, ISSN: 0146-1672.

[44] R. J. Larsen and M. Kasimatis, "Individual differences in entrainment of mood to the weekly calendar," eng, *Journal of personality and social psychology*, vol. 58, no. 1, pp. 164–171, 1990, ISSN: 0022-3514.

[45] S. Reid, A. Towell, and J. Golding, "Seasonality, social zeitgebers and mood variability in entrainment of mood: Implications for seasonal affective disorder," eng, *Journal of affective disorders*, vol. 59, no. 1, pp. 47–54, 2000, ISSN: 0165-0327.

[46] H. T. Reis, K. M. Sheldon, S. L. Gable, J. Roscoe, and R. M. Ryan, "Daily well-being: The role of autonomy, competence, and relatedness," eng, *Personality & social psychology bulletin*, vol. 26, no. 4, pp. 419–435, 2000, ISSN: 0146-1672.

[47] R. M. Ryan, J. H. Bernstein, and K. W. Brown, "Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms," eng, *Journal of social and clinical psychology*, vol. 29, no. 1, pp. 95–122, 2010, ISSN: 0736-7236.

[48] M. F. Taylor, M. E. Edwards, and J. A. Pooley, ""nudging them back to reality": Toward a growing public acceptance of the role dogs fulfill in ameliorating contemporary veterans' ptsd symptoms," eng, *Anthrozoös*, vol. 26, no. 4, pp. 593–611, 2013, ISSN: 0892-7936.

[49] D. Scotland-Coogan, *Receiving and training a service dog: The impact on combat veterans with posttraumatic stress disorder (ptsd)*, eng, 2017.

[50] B. J. H. Yarborough, S. P. Stumbo, M. T. Yarborough, A. Owen-Smith, and C. A. Green, "Benefits and challenges of using service dogs for veterans with posttraumatic stress disorder," eng, *Psychiatric rehabilitation journal*, vol. 41, no. 2, pp. 118–124, 2018, ISSN: 1095-158X.

[51] M. L. Kloep, R. H. Hunter, and S. J. Kertz, "Examining the effects of a novel training program and use of psychiatric service dogs for military-related ptsd and associated symptoms," eng, *American journal of orthopsychiatry*, vol. 87, no. 4, pp. 425–433, 2017, ISSN: 0002-9432.

[52] N. G. Polson, J. G. Scott, and J. Windle, "Bayesian inference for logistic models using pólya-gamma latent variables," eng, *Journal of the American Statistical Association*, vol. 108, no. 504, pp. 1339–1349, 2013, ISSN: 0162-1459. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/01621459.2013.829001.

[53] C. E. Frangakis and D. B. Rubin, "Principal stratification in causal inference," *Biometrics*, vol. 58, no. 1, pp. 21–29, 2002, ISSN: 0006-341X.

[54] C. E. Frangakis, R. S. Brookmeyer, R. Varadhan, M. Safaeian, D. Vlahov, and S. A. Strathdee, "Methodology for evaluating a partially controlled longitudinal treatment using principal stratification, with application to a needle exchange program," eng, *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 239–249, 2004, ISSN: 0162-1459.

[55] C.-P. Wang, B. Jo, and C. Hendricks Brown, "Causal inference in longitudinal comparative effectiveness studies with repeated measures of a continuous intermediate variable," eng, *Statistics in medicine*, vol. 33, no. 20, pp. 3509–3527, 2014, ISSN: 0277-6715.

[56] B. Bornkamp, K. Rufibach, J. Lin, Y. Liu, D. V. Mehrotra, S. Roychoudhury, H. Schmidli, Y. Shentu, and M. Wolbers, "Principal stratum strategy: Potential role in drug development," eng, *Pharmaceutical statistics : the journal of the pharmaceutical industry*, 2021, ISSN: 1539-1604.

[57] A. Sarkar and D. B. Dunson, "Bayesian nonparametric modeling of higher order markov chains," eng, *Journal of the American Statistical Association*, vol. 111, no. 516, pp. 1791–1803, 2016, ISSN: 0162-1459.

[58] F. B. Hu, J. Goldberg, D. Hedeker, B. R. Flay, and M. A. Pentz, "Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes," eng, *American journal of epidemiology*, vol. 147, no. 7, pp. 694–703, 1998, ISSN: 0002-9262.

[59] P. J. Heagerty and S. L. Zeger, "Marginalized multilevel models and likelihood inference," eng, *Statistical science*, vol. 15, no. 1, pp. 1–19, 2000, ISSN: 0883-4237.

[60] J. K. Lindsey and P. Lambert, "On the appropriateness of marginal models for repeated measurements in clinical trials," eng, *Statistics in medicine*, vol. 17, no. 4, pp. 447–469, 1998, ISSN: 0277-6715.

[61] *Analysis of longitudinal data.* eng, 2nd ed. / Peter J. Diggle ... [et al.]., ser. Oxford statistical science series ; 25. Oxford ; New York: Oxford University Press, 2002, ISBN: 0198524846.

[62] J. E. Overall and S. Tonidandel, "Robustness of generalized estimating equation (gee) tests of significance against misspecification of the error structure model," eng, *Biometrical journal*, vol. 46, no. 2, pp. 203–213, 2004, ISSN: 0323-3847.

[63] J. W. (W. Hardin, *Generalized estimating equations*, eng, Second edition. 2013, ISBN: 9781439881132.

[64] N. E. Breslow and D. G. Clayton, "Approximate inference in generalized linear mixed models," eng, *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 9–25, 1993, ISSN: 0162-1459.

[65] F. (Liang, *Advanced Markov chain Monte Carlo methods : learning from past samples*, eng, ser. Wiley series in computational statistics. Chichester, West Sussex, U.K.: Wiley, 2010, ISBN: 9780470748268.

[66] W. Pan, "Akaikes information criterion in generalized estimating equations," eng, *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001, ISSN: 0006-341X.

[67] N. Koper and M. Manseau, "Generalized estimating equations and generalized linear mixed-effects models for modelling resource selection," eng, *The Journal of applied ecology*, vol. 46, no. 3, pp. 590–599, 2009, ISSN: 0021-8901.

[68] N. G. Polson, J. G. Scott, and J. Windle, "Bayesian inference for logistic models using pólya–gamma latent variables," *Journal of the American Statistical Association*, vol. 108, no. 504, pp. 1339–1349, 2013. DOI: 10.1080/01621459.2013.829001. eprint: https://doi.org/10.1080/01621459.2013.829001. [Online]. Available: https://doi.org/10.1080/01621459.2013.829001.

[69] R. M. Neal, "Mcmc using hamiltonian dynamics," eng, 2012.

[70] D. Gamerman, "Sampling from the posterior distribution in generalized linear mixed models," eng, *Statistics and computing*, vol. 7, no. 1, pp. 57–68, 1997, ISSN: 0960-3174.

[71] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," eng, *The Journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953, ISSN: 0021-9606.

[72] J. Chen and J. Shao, "Iterative weighted least squares estimators," eng, *The Annals of statistics*, vol. 21, no. 2, pp. 1071–1092, 1993, ISSN: 0090-5364.

[73] P. P. McCullagh, *Generalized linear models*, eng, Second edition., ser. Monographs on statistics and applied probability Generalized linear models ; 37. 1989, ISBN: 0-412-31760-5.

[74] C. C. Holmes and L. Held, "Bayesian auxiliary variable models for binary and multinomial regression," eng, *Bayesian analysis*, vol. 1, no. 1 A, pp. 145–168, 2006, ISSN: 1936-0975.

[75] S. Frühwirth-Schnatter and R. Frühwirth, "Data augmentation and mcmc for binary and multinomial logit models," eng, in *Statistical Modelling and Regression Structures*, Heidelberg: Physica-Verlag HD, 2009, pp. 111–132, ISBN: 3790824127.

[76] R. B. Gramacy and N. G. Polson, "Simulation-based regularized logistic regression," eng, *Bayesian analysis*, vol. 7, no. 3, pp. 567–590, 2012, ISSN: 1936-0975.

[77] J. W. Pillow and J. Scott, "Fully bayesian inference for neural models with negative-binomial spiking," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012, pp. 1898–1906. [Online]. Available: http://papers.nips.cc/paper/4567-fully-bayesian-inference-for-neural-models-with-negative-binomial-spiking.pdf.

[78] J. H. Kook, K. A. Vaughn, D. M. DeMaster, L. Ewing-Cobbs, and M. Vannucci, "Bvar-connect: A variational bayes approach to multi-subject vector autoregressive models for inference on brain connectivity networks," eng, *Neuroinformatics (Totowa, N.J.)*, vol. 19, no. 1, pp. 39–56, 2021, ISSN: 1539-2791.

[79] C. Koki, L. Meligkotsidou, and I. Vrontos, "Forecasting under model uncertainty: Non-homogeneous hidden markov models with pòlya-gamma data augmentation," eng, *Journal of forecasting*, vol. 39, no. 4, pp. 580–598, 2020, ISSN: 0277-6693.

[80] T. Krisztin and P. Piribauer, "A bayesian spatial autoregressive logit model with an empirical application to european regional fdi flows," eng, *Empirical economics*, 2020, ISSN: 0377-7332.

[81] M. D. Hoffman and A. Gelman, "The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo," eng, 2011.

[82] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper)," eng, *Bayesian analysis*, vol. 1, no. 3, pp. 515–534, 2006, ISSN: 1936-0975.

[83] E. Hariton and J. J. Locascio, "Randomised controlled trials - the gold standard for effectiveness research: Study design: Randomised controlled trials," eng, *BJOG : an international journal of obstetrics and gynaecology*, vol. 125, no. 13, pp. 1716–1716, 2018, ISSN: 1470-0328.

[84] M. Akacha, F. Bretz, D. Ohlssen, G. Rosenkranz, and H. Schmidli, "Estimands and their role in clinical trials," eng, *Statistics in biopharmaceutical research*, vol. 9, no. 3, pp. 268–271, 2017, ISSN: 1946-6315.

[85] J. R. Carpenter, J. H. Roger, S. Cro, and M. G. Kenward, "Response to comments by seaman et al. on "analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation," journal of biopharmaceutical statistics 23:1352-1371," eng, *Journal of biopharmaceutical statistics*, vol. 24, no. 6, pp. 1363–1369, 2014, ISSN: 1054-3406.

[86] *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials*, eng, 90. Washington: Federal Information & News Dispatch, LLC, 2021, vol. 86, p. 26 047.

[87] Y. Qu, J. Luo, and S. J. Ruberg, "Implementation of tripartite estimands using adherence causal estimators under the causal inference framework," eng, *Pharmaceutical statistics : the journal of the pharmaceutical industry*, vol. 20, no. 1, pp. 55–67, 2021, ISSN: 1539-1604.

[88] M. Akacha, F. Bretz, and S. Ruberg, "Estimands in clinical trials – broadening the perspective," *Statistics in Medicine*, vol. 36, no. 1, pp. 5–19, 2017, ISSN: 0277-6715.

[89] O. Sofrygin, M. J. van der Laan, and R. Neugebauer, "Simcausal r package: Conducting transparent and reproducible simulation studies of causal effect estimation with complex longitudinal data," eng, *Journal of statistical software*, vol. 81, no. 2, pp. 1–47, 2017, ISSN: 1548-7660.

[90] J. Pearl, "Causal diagrams for empirical research," eng, *Biometrika*, vol. 82, no. 4, pp. 669–688, 1995, ISSN: 0006-3444.

[91] G. Paux and A. Dmitrieniko, *Mediana: An r package for clinical trial simulations*, 2019.

[92] N. Benda, M. Branson, W. Maurer, and T. Friede, "Aspects of modernizing drug development using clinical scenario planning and evaluation," eng, *Therapeutic innovation & regulatory science*, vol. 44, no. 3, pp. 299–315, 2010, ISSN: 2168-4790.

[93] T. Friede, R. Nicholas, N. Stallard, S. Todd, N. Parsons, E. Valdés-Márquez, and J. Chataway, "Refinement of the clinical scenario evaluation framework for assessment of competing development strategies with an application to multiple sclerosis," eng, *Therapeutic innovation & regulatory science*, vol. 44, no. 6, pp. 713–718, 2010, ISSN: 2168-4790.

[94] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," eng, *Journal of educational psychology*, vol. 66, no. 5, pp. 688–701, 1974, ISSN: 0022-0663.

[95] K. Thorlund, S. Golchi, J. Haggstrom, and E. Mills, "Highly efficient clinical trials simulator (hect): Software application for planning and simulating platform adaptive trials," eng, *Gates open research*, vol. 3, pp. 780–780, 2019, ISSN: 2572-4754.

[96] J. Wojciechowski, A. Hopkins, and R. Upton, "Interactive pharmacometric applications using r and the shiny package: Interactive pharmacometric applications with shiny," eng, *CPT: pharmacometrics and systems pharmacology*, vol. 4, no. 3, pp. 146–159, 2015, ISSN: 2163-8306.

[97] M. J. Grayling and J. M. Wason, "A web application for the design of multi-arm clinical trials," eng, *BMC cancer*, vol. 20, no. 1, pp. 80–80, 2020, ISSN: 1471-2407.

[98] A. Karanevich, R. Meier, S. Graw, A. McGlothlin, and B. Gajewski, "Optimizing sample size allocation and power in a bayesian two-stage drop-the-losers design," eng, *The American statistician*, vol. 75, no. 1, pp. 66–75, 2021, ISSN: 0003-1305.

[99] B. Karges, J. Rosenbauer, T. Kapellen, V. M. Wagner, E. Schober, W. Karges, and R. W. Holl, "Hemoglobin a1c levels and risk of severe hypoglycemia in children and young adults with type 1 diabetes from germany and austria: A trend analysis in a cohort of 37,539 patients between 1995 and 2012," eng, *PLoS medicine*, vol. 11, no. 10, e1001742–e1001742, 2014, ISSN: 1549-1277.

[100] G. Verdile, S. Fuller, C. S. Atwood, S. M. Laws, S. E. Gandy, and R. N. Martins, "The role of beta amyloid in alzheimer's disease: Still a cause of everything or the only one who got caught?" eng, *Pharmacological research*, vol. 50, no. 4, pp. 397–409, 2004, ISSN: 1043-6618.

[101] K. Stenlöf, W. T. Cefalu, K.-A. Kim, M. Alba, K. Usiskin, C. Tong, W. Canovatchel, and G. Meininger, "Efficacy and safety of canagliflozin monotherapy in subjects with type 2 diabetes mellitus inadequately controlled with diet and exercise," eng, *Diabetes, obesity & metabolism*, vol. 15, no. 4, pp. 372–382, 2013, ISSN: 1462-8902.

[102] M. A. Mintun, A. C. Lo, C. Duggan Evans, A. M. Wessels, P. A. Ardayfio, S. W. Andersen, S. Shcherbinin, J. Sparks, J. R. Sims, M. Brys, L. G. Apostolova, S. P. Salloway, and D. M. Skovronsky, "Donanemab in early alzheimer's disease," eng, *The New England journal of medicine*, vol. 384, no. 18, pp. 1691–1704, 2021, ISSN: 0028-4793.

[103] Z. Jiang and J. Templin, "Gibbs samplers for logistic item response models via the pólya–gamma distribution: A computationally efficient data-augmentation strategy," eng, *Psychometrika*, vol. 84, no. 2, pp. 358–374, 2019, ISSN: 0033-3123.

[104] J. Pillow and J. Scott, "Fully bayesian inference for neural models with negative-binomial spiking," vol. 3, 2012, pp. 1898–1906, ISBN: 9781627480031.

A. SUPPLEMENTARY MATERIAL FOR CHAPTER 1

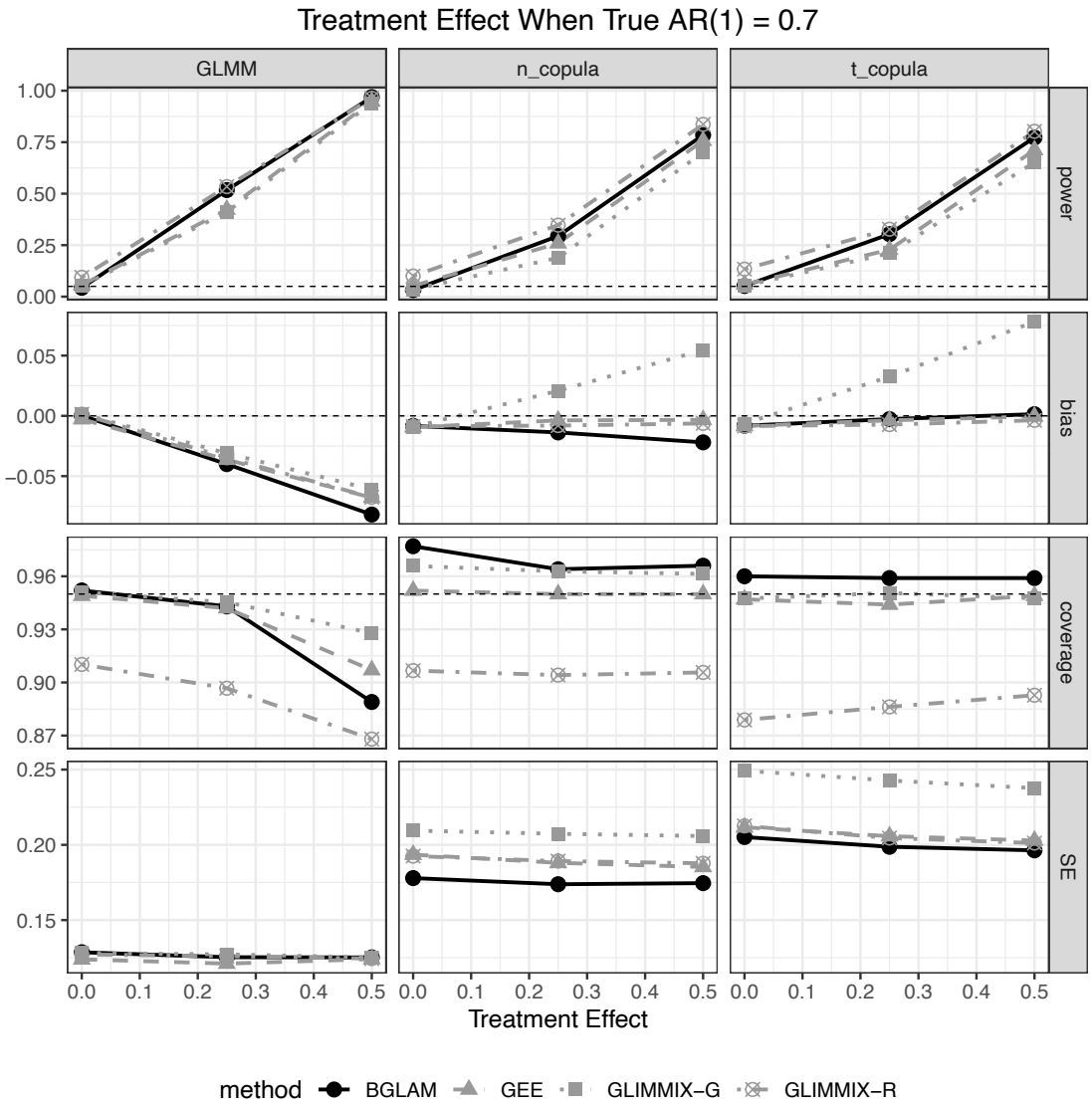


Figure A.1. Plots of power, bias, coverage and standard errors of β_{trt} when AR(1)=0.7 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has high power, low bias, good coverage and low standard error, except when the data were generated via GLMM at $\beta_{\text{trt}} = 0.5$. This is expected, since there is a mismatch between the model and the data generated.

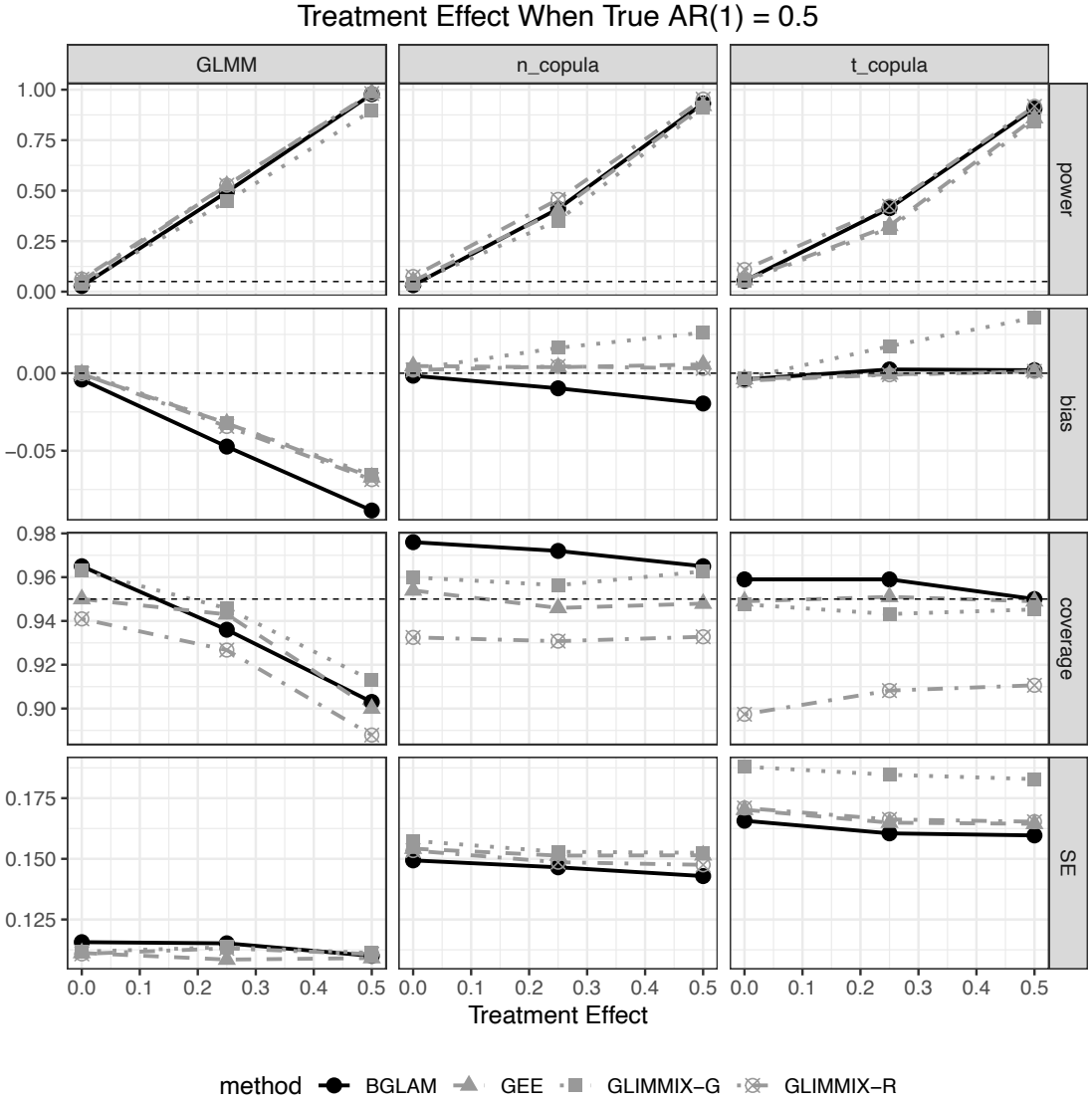


Figure A.2. Plots of power, bias, coverage and standard errors of β_{trt} when AR(1)=0.5 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has high power, low bias, good coverage and low standard error, except when the data were generated via GLMM at $\beta_{\text{trt}} = 0.5$. This is expected, since there is a mismatch between the model and the data generated.

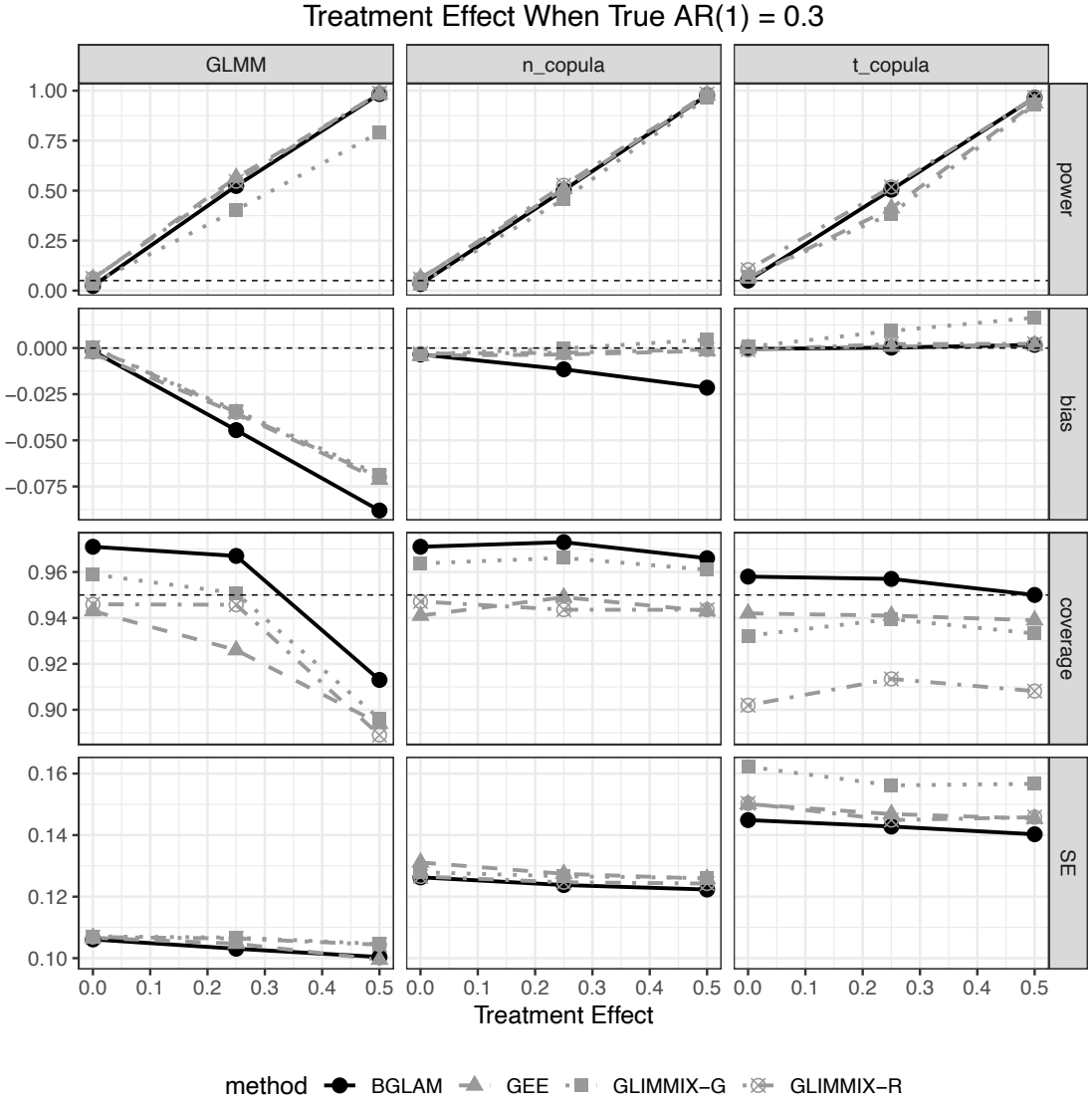


Figure A.3. Plots of power, bias, coverage and standard errors of β_{trt} when AR(1)=0.3 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has high power, low bias, good coverage and low standard error, except when the data were generated via GLMM at $\beta_{\text{trt}} = 0.5$. This is expected, since there is a mismatch between the model and the data generated.

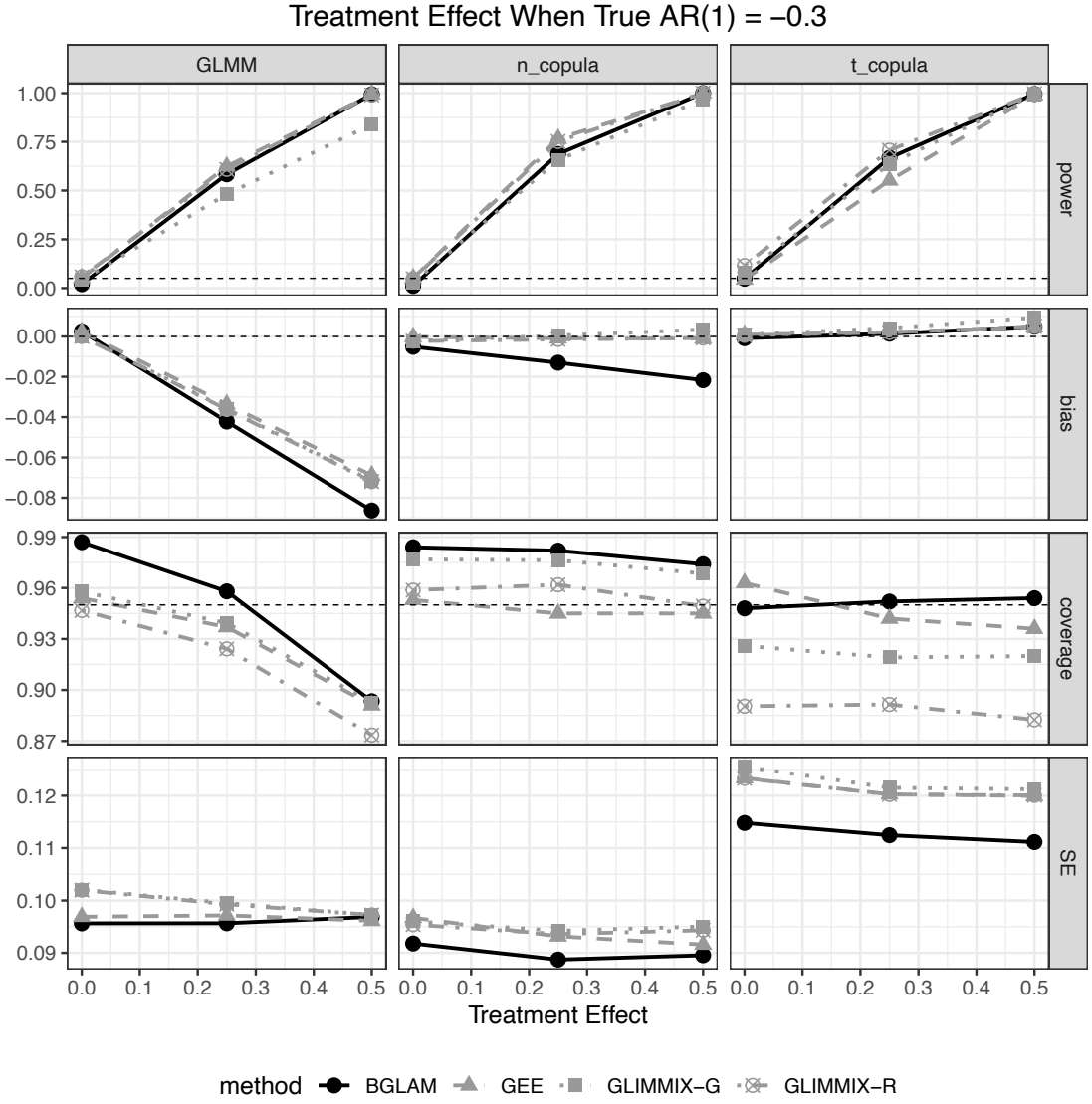


Figure A.4. Plots of power, bias, coverage and standard errors of β_{trt} when AR(1)=-0.3 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has high power, low bias, good coverage and low standard error, except when the data were generated via GLMM at $\beta_{\text{trt}} = 0.5$. This is expected, since there is a mismatch between the model and the data generated.

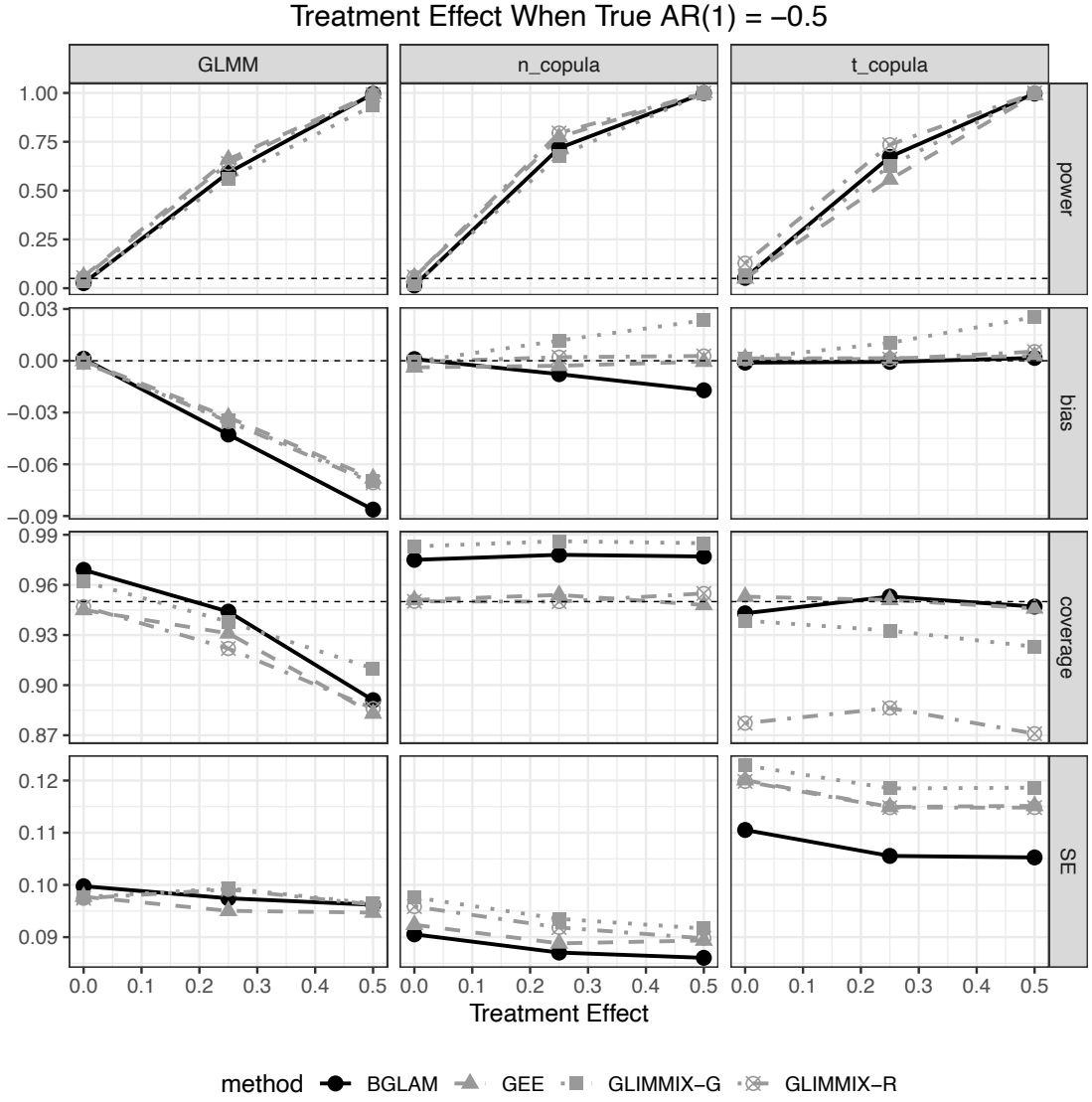


Figure A.5. Plots of power, bias, coverage and standard errors of β_{trt} when AR(1)=-0.5 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has high power, low bias, good coverage and low standard error, except when the data were generated via GLMM at $\beta_{\text{trt}} = 0.5$. This is expected, since there is a mismatch between the model and the data generated. Although GLIMMIX-G and GLIMMIX-R have high power, the elevate Type I error rates average around 0.9 at $\alpha = 0.05$.

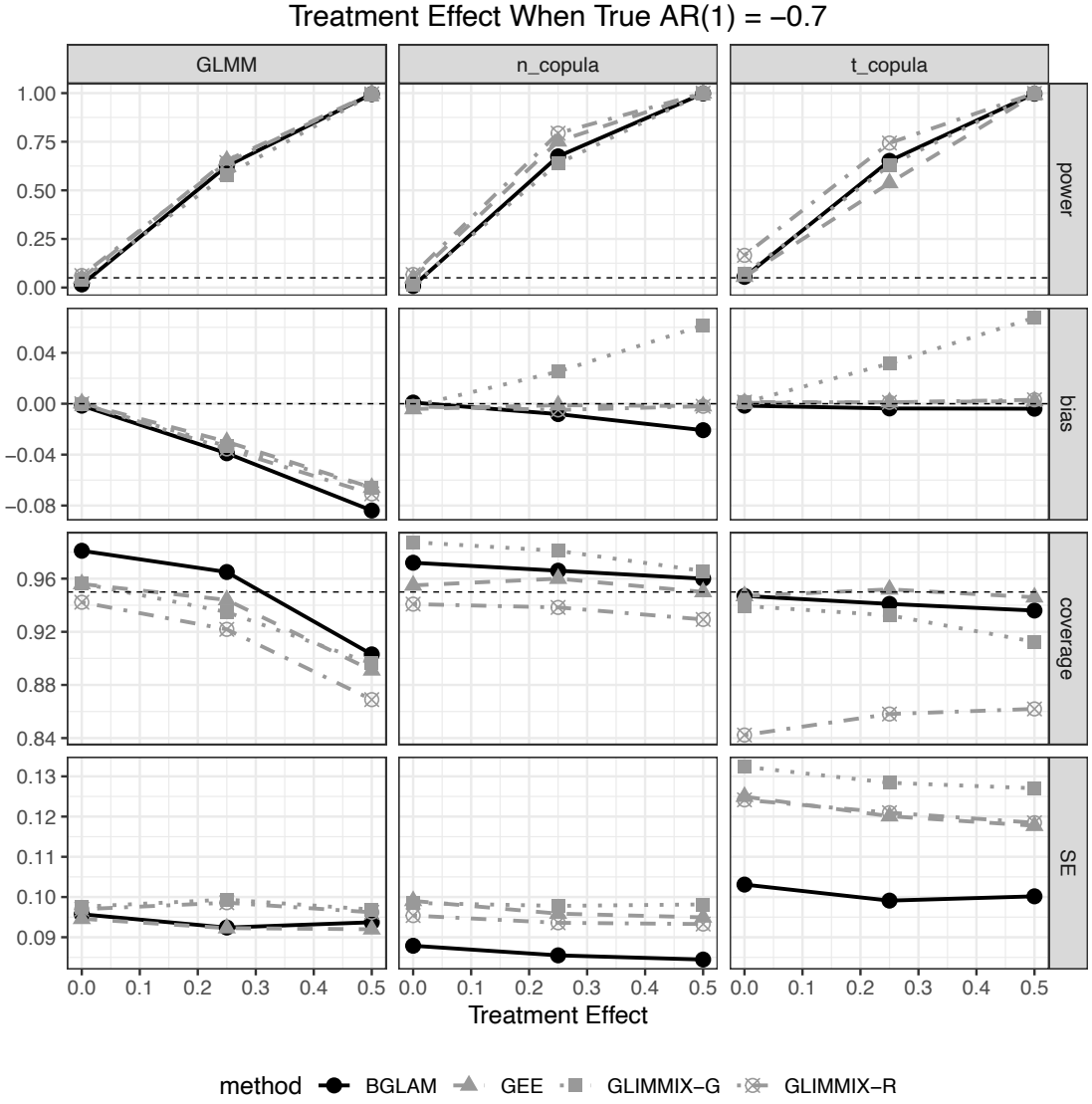


Figure A.6. Plots of power, bias, coverage and standard errors of β_{trt} when AR(1)=-0.7 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has high power, low bias, good coverage and low standard error, except when the data were generated via GLMM at $\beta_{\text{trt}} = 0.5$. This is expected, since there is a mismatch between the model and the data generated. Although GLIMMIX-G and GLIMMIX-R have high power, the elevate Type I error rates average around 0.9 at $\alpha = 0.05$.

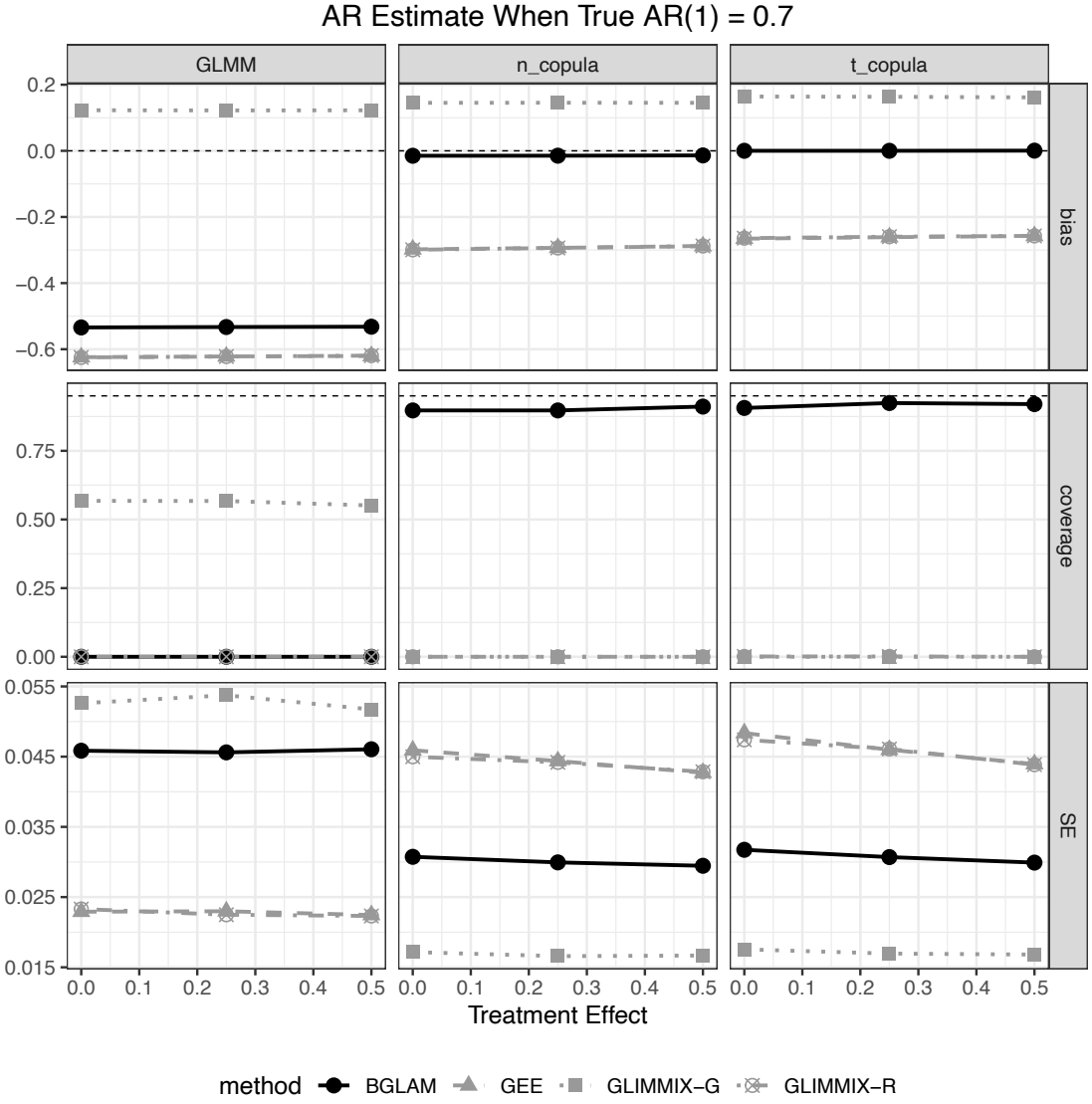


Figure A.7. Plots of bias, coverage and standard errors of the AR(1) estimate when the true AR(1)=0.7 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has low bias and standard errors with good coverage, with the exception when data were generated via GLMM. This is expected, since there is a mismatch between the model and the data generated.

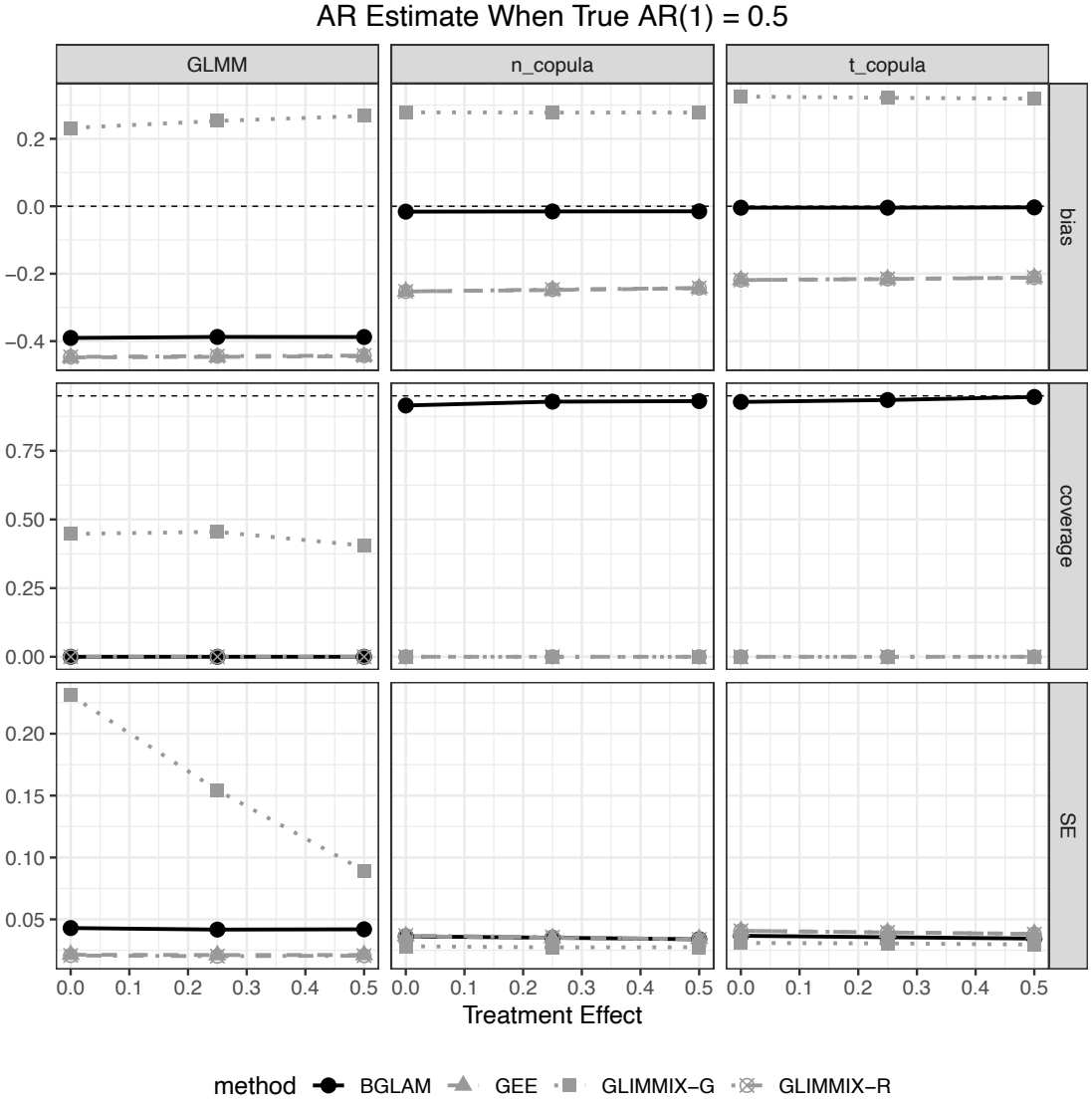


Figure A.8. Plots of bias, coverage and standard errors of the AR(1) estimate when the true AR(1)=0.5 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has low bias and standard errors with good coverage, with the exception when data were generated via GLMM. This is expected, since there is a mismatch between the model and the data generated.

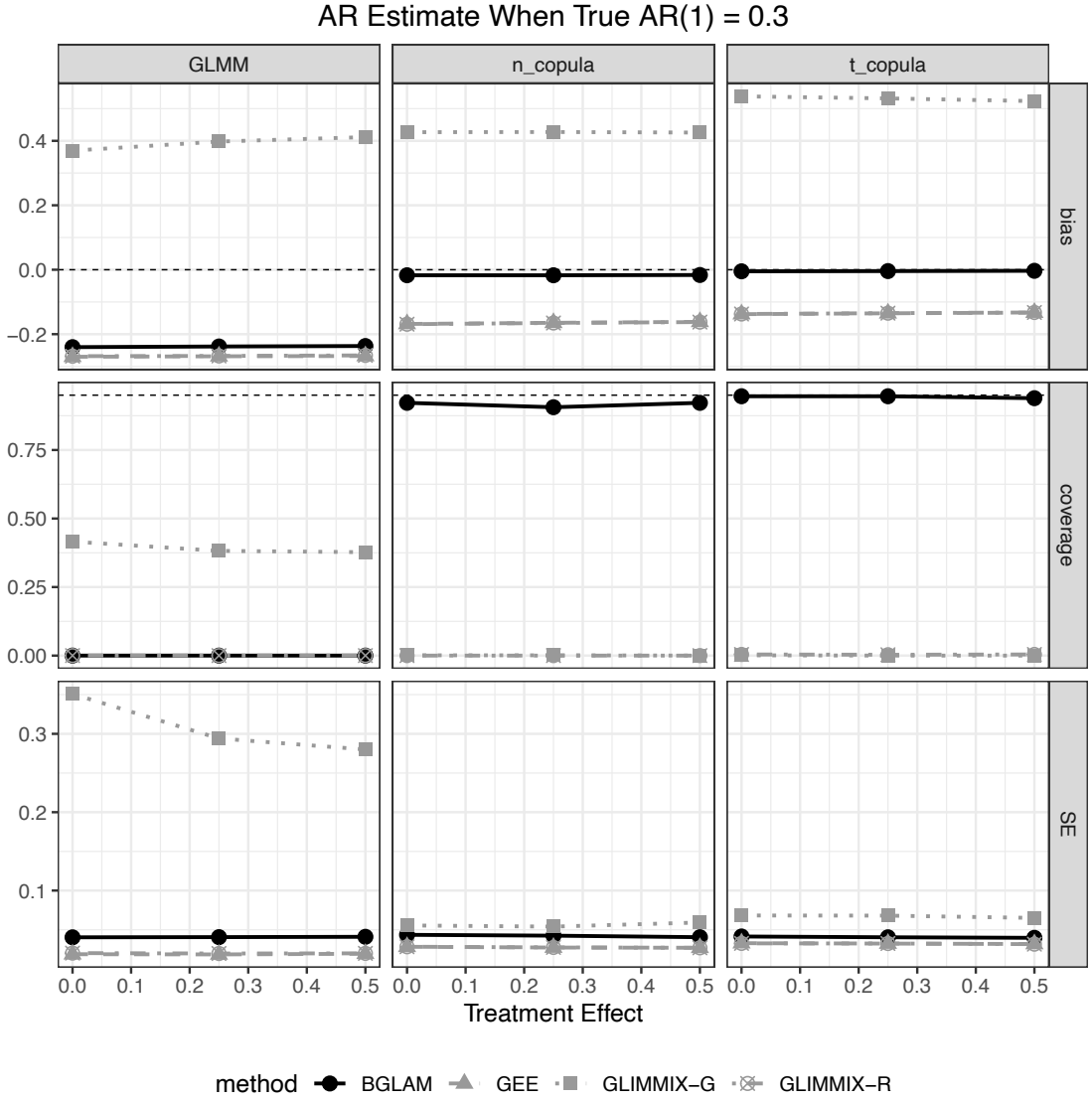


Figure A.9. Plots of bias, coverage and standard errors of the AR(1) estimate when the true AR(1)=0.3 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has low bias and standard errors with good coverage, with the exception when data were generated via GLMM. This is expected, since there is a mismatch between the model and the data generated.

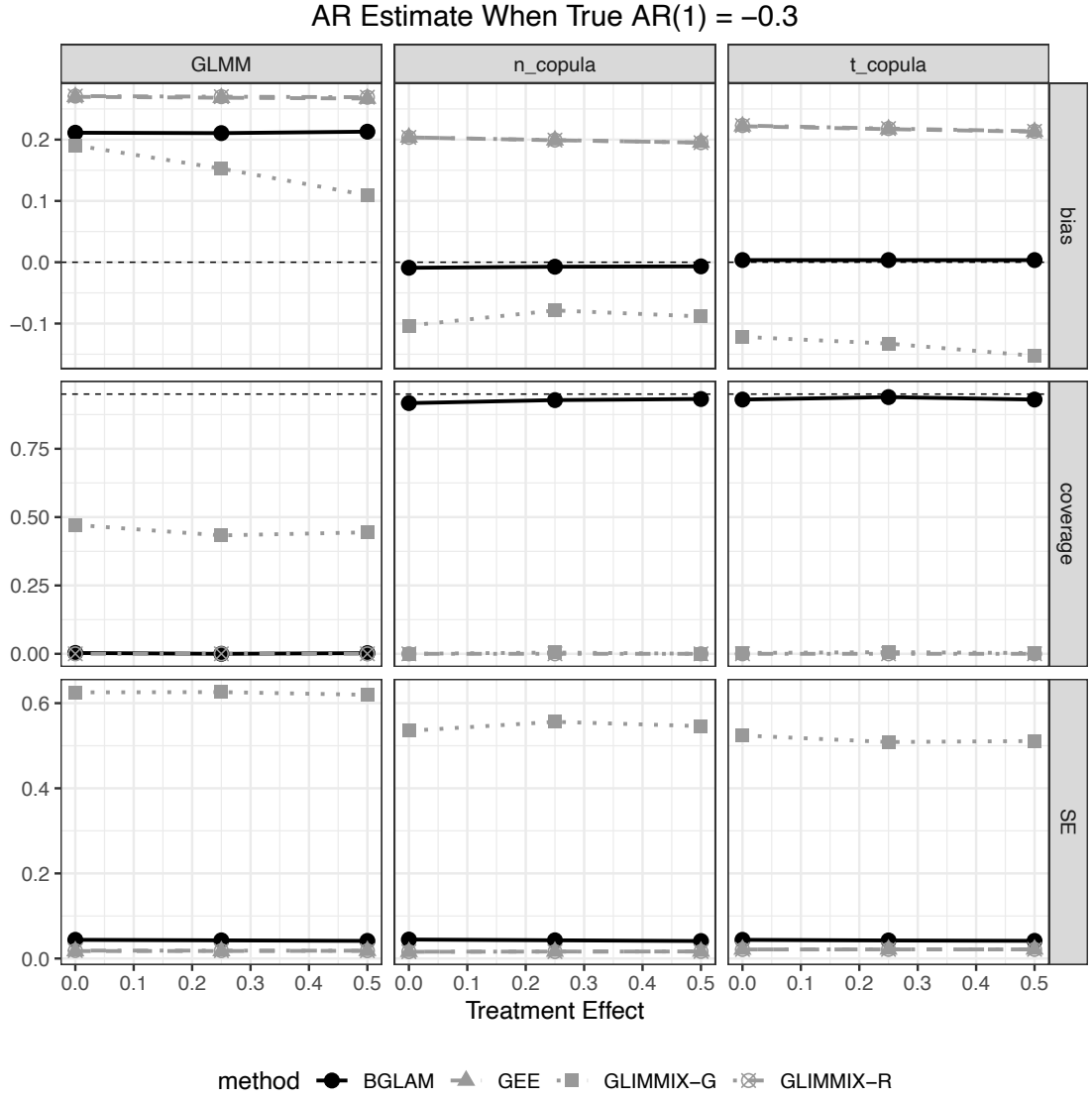


Figure A.10. Plots of bias, coverage and standard errors of the AR(1) estimate when the true AR(1)=-0.3 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has low bias and standard errors with good coverage, with the exception when data were generated via GLMM. This is expected, since there is a mismatch between the model and the data generated.

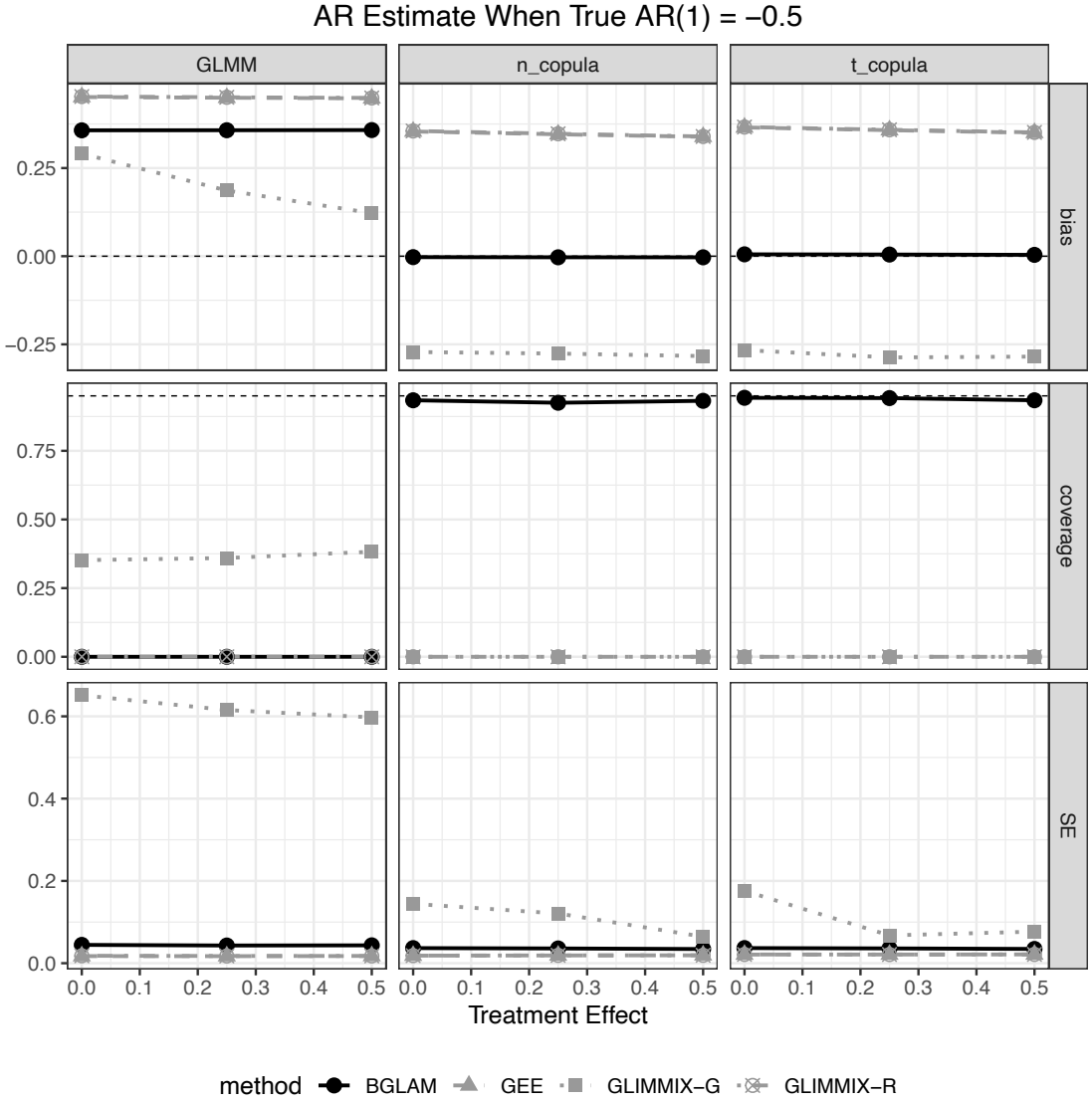


Figure A.11. Plots of bias, coverage and standard errors of the AR(1) estimate when the true AR(1)=-0.5 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has low bias and standard errors with good coverage, with the exception when data were generated via GLMM. This is expected, since there is a mismatch between the model and the data generated.

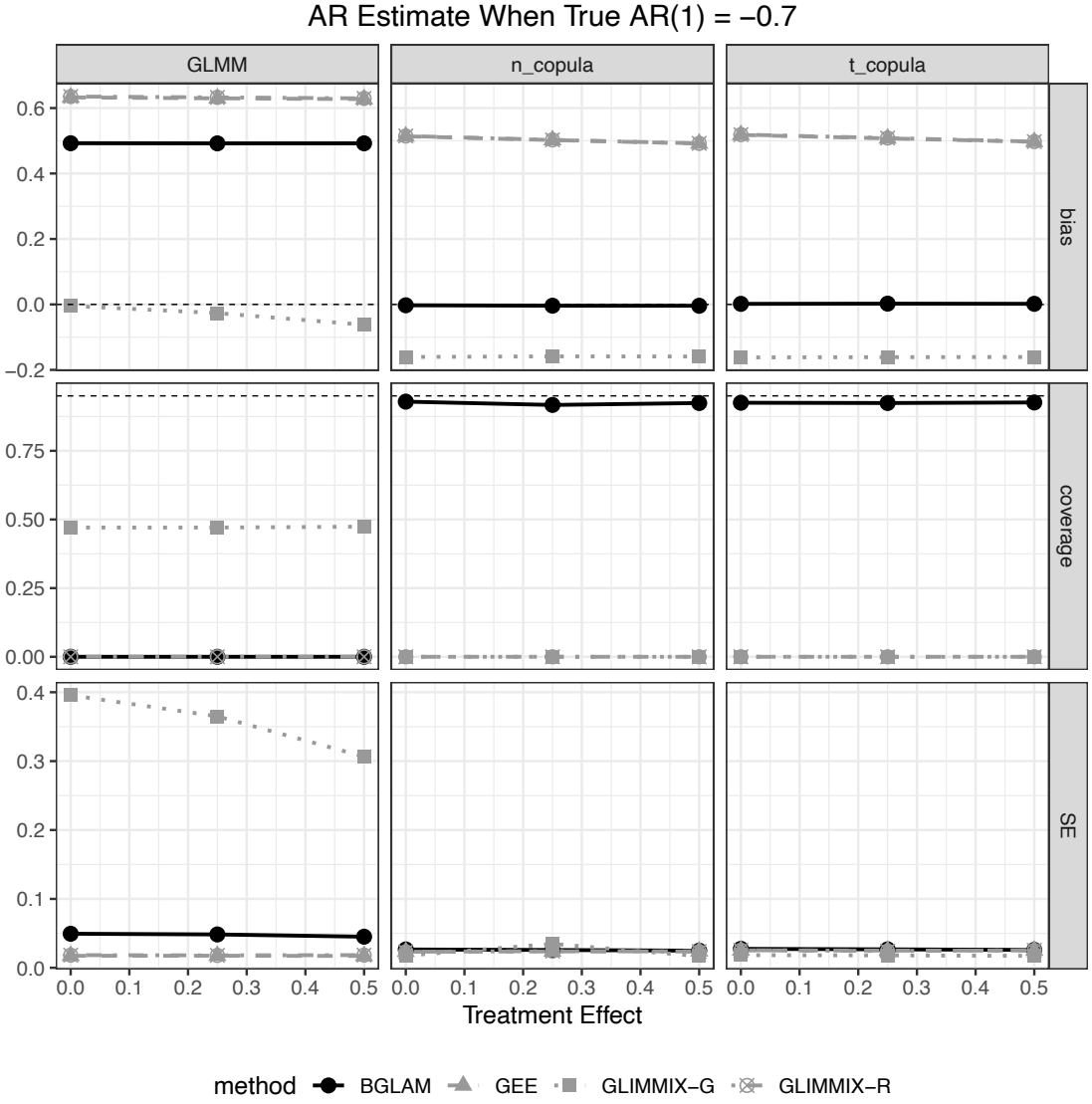


Figure A.12. Plots of bias, coverage and standard errors of the AR(1) estimate when the true AR(1)=-0.7 for 1000 simulated datasets, with the BGLAM being the solid black line. BGLAM has low bias and standard errors with good coverage, with the exception when data were generated via GLMM. This is expected, since there is a mismatch between the model and the data generated.

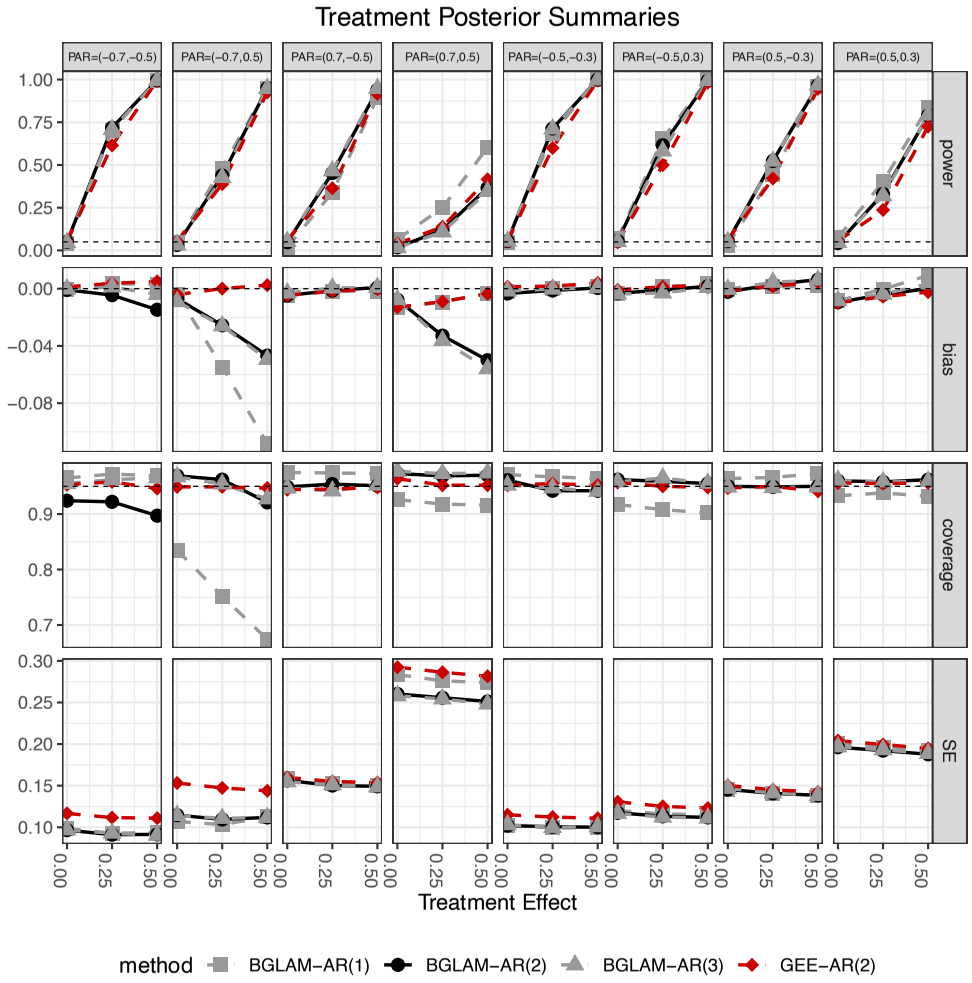


Figure A.13. Plots of power, coverage, bias and standard errors for 1000 simulated datasets, with the correct BGLAM-AR(2) model being the solid black line. The BGLAM-AR(2) model has consistently higher power and lower SE than GEE-AR(2) (dashed red line). Although all BGLAM models are slightly biased when AR(2)=(-0.7, 0.5), the BGLAM-AR(1) model is twice as biased and suffers from undercoverage. The exception is when AR(s)=(0.7, 0.5), where BGLAM-AR(1) has the least amount of bias. Although BGLAM-AR(1) has the highest power in this simulation setting, it is still low in comparison to the power under different simulation settings.

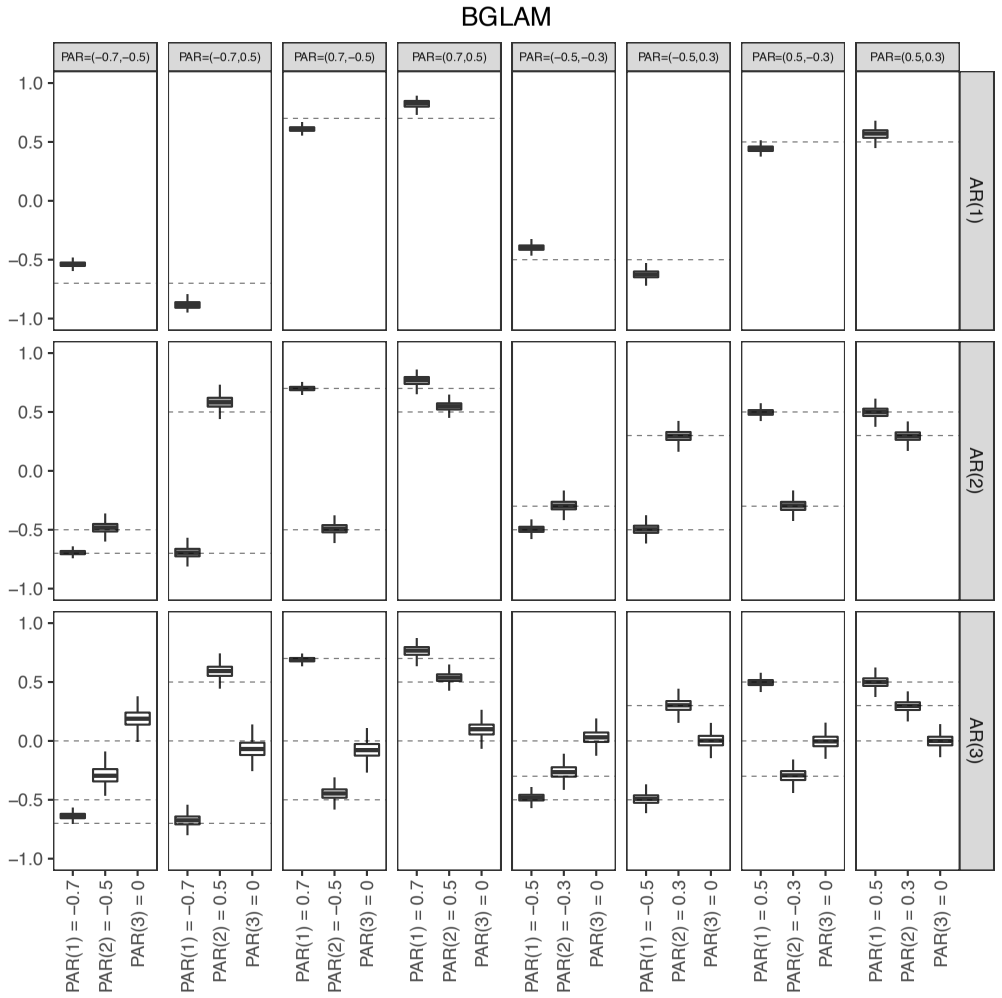


Figure A.14. Boxplots of the PARs for 1000 simulated datasets when $\beta_{\text{trt}} = 0$, with the true correct model of AR(2) being the filled grey boxplots in the middle row. The horizontal dotted lines are the true PAR settings. The boxplots of the PARs in the AR(2) model (second row) cover the true values much better than the AR(1) model (first row). The boxplots of the last PAR for the AR(3) model is centered around 0 while the boxplots of the first and second PARs of the same model are centered around the true input parameters (third row). Results are consistent across the other β_{trt} values.

Table A.1. Posterior summaries of the varying PARs under different AR(k) settings used to supplement model selection. The 95% CIs of the last PARs for AR(3) for the weekday and weekend partitions cover zero (highlighted grey cells for AR(3)) while none of the 95% CI for the first and second PARs covers zero. Favoring parsimony, this compels us to model the weekday correlation structure with an AR(2). Further, 95% CI for the first PAR in the weekend partition covers zero (highlighted grey cells for AR(2)). In combination with the DIC and Marginal LL, we decided on an AR(2) correlation structure for weekdays and an AR(1) correlation structure for weekends (bottom highlighted grey cells).

AR(k)	DIC	Marginal LL	Partition	k	PAR Posterior Mean	95% Lower CI	95% Upper CI	β_{trt}	ppp_{IS}
1	117.09	-70.98	Weekday	1	0.28 (0.04)	0.19	0.37	0.1 (0.09)	0.09
			Thur-Fri	NA	0.03 (0.16)	-0.28	0.34		
			Weekend	1	0.13 (0.27)	-0.01	0.27		
			Sun-Mon	NA	0.07 (0.12)	-0.18	0.34		
2	117.41	-70.94	Weekday	1	0.20 (0.05)	0.1	0.3	0.04 (0.1)	0.13
				2	0.41 (0.04)	0.33	0.5		
			Thur-Fri	NA	0.14 (0.11)	-0.09	0.56		
			Weekend	1	0.09 (0.06)	-0.03	0.21		
				2	0.37 (0.05)	0.27	0.48		
			Sun-Mon	NA	0.15 (0.90)	-0.03	0.33		
3	117.52	-70.88	Weekday	1	0.19 (0.04)	0.1	0.27	0.06 (0.12)	0.17
				2	0.41 (0.03)	0.34	0.47		
				3	0.03 (0.06)	-0.1	0.16		
			Thur-Fri	NA	0.09 (0.14)	-0.2	0.38		
			Weekend	1	0.09 (0.09)	-0.09	0.28		
				2	0.39 (0.05)	0.3	0.48		
				3	-0.09 (0.07)	-0.24	0.06		
			Sun-Mon	NA	0.17 (0.09)	-0.01	0.35		
2+1	**116.28**	**-70.75**	**Weekday**	**1**	**0.21 (0.06)**	**0.09**	**0.33**	**0.06 (0.09)**	**0.08**
				2	**0.4 (0.04)**	**0.32**	**0.48**		
			Thur-Fri	**NA**	**0.1 (0.16)**	**0.02**	**0.25**		
			Weekend	**1**	**0.14 (0.06)**	**0.02**	**0.25**		
			Sun-Mon	**NA**	**0.1 (0.12)**	**-0.14**	**0.35**		

A.A SAS Script

```
proc glimmix data= final;
class y treatment(ref="0") subject time;
model y(event = "1") = treatment x_0 / ddfm=kr dist=binary link=logit solution;
random time / type=ar(1) subject=subject;
by seed_val;
run;

proc glimmix data= final;
class y treatment(ref="0") subject time;
model y(event = "1") = treatment x_0 / ddfm=kr dist=binary link=logit solution;
random time / type=ar(1) subject=subject residual;
by seed_val;
run;
```

B. SUPPLEMENTARY MATERIAL FOR CHAPTER 2

B.A Simulation Results

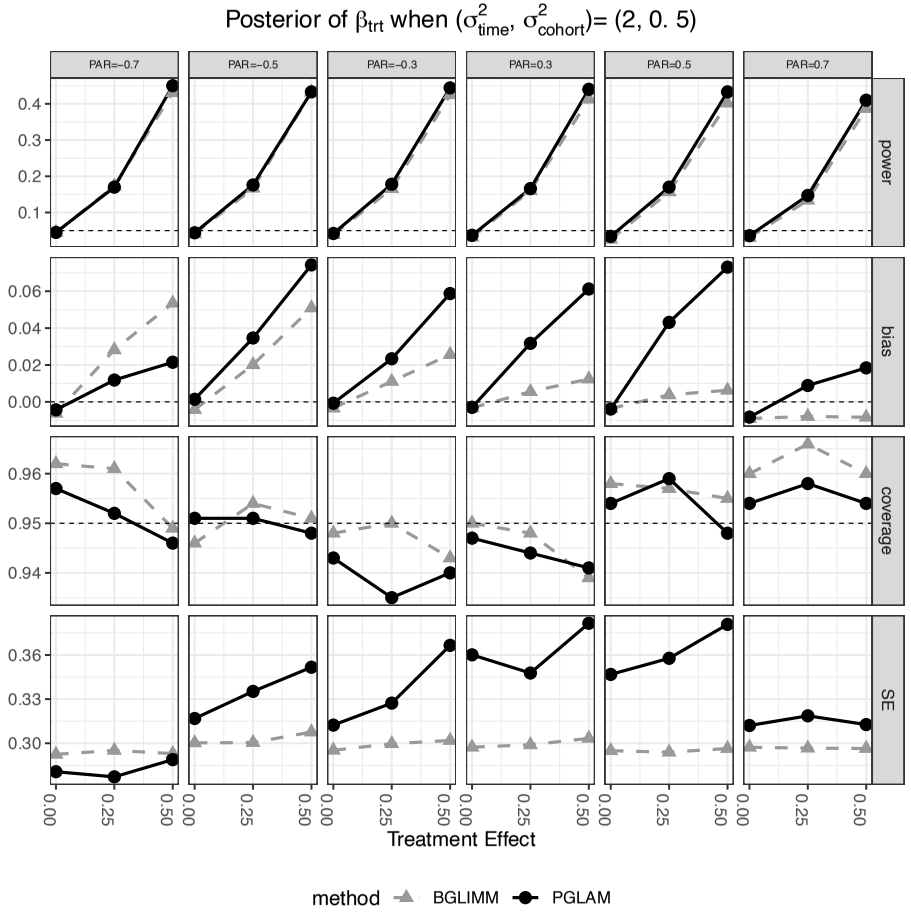


Figure B.1. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 0.5)$. Although both models have similar power and coverage, GLAMRE has higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

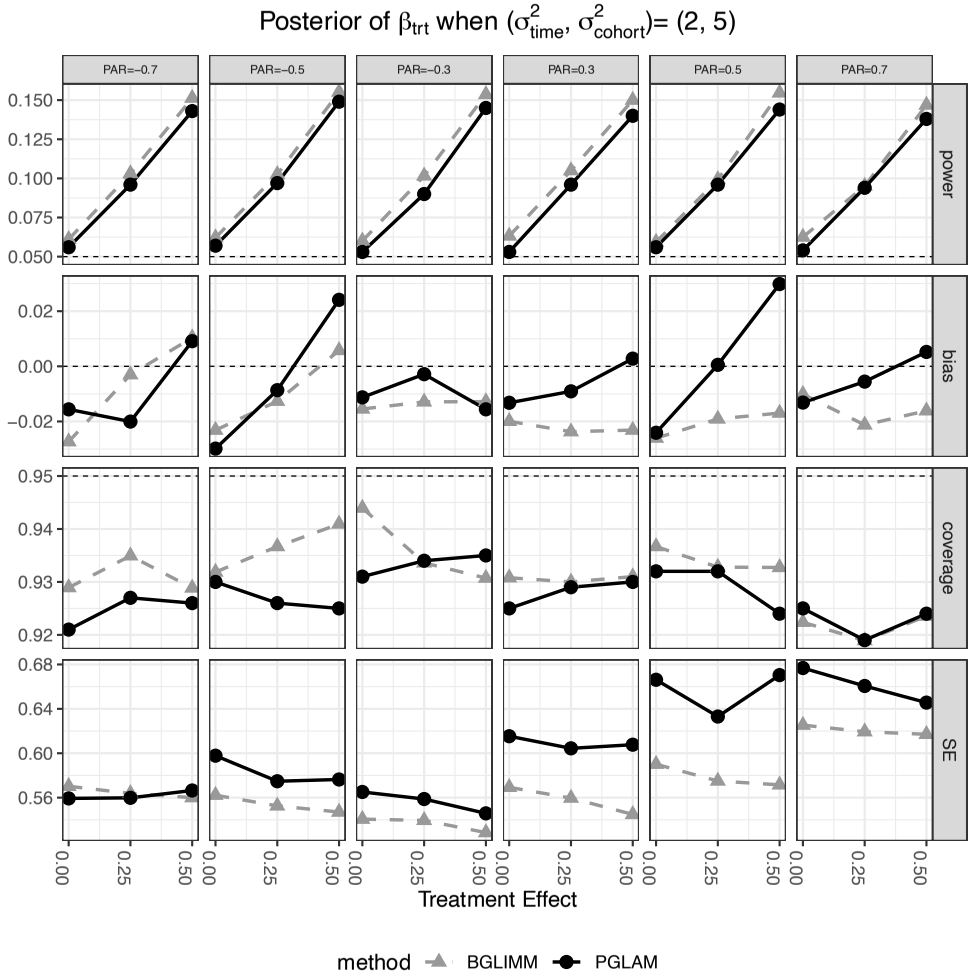


Figure B.2. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 5)$. Although both models have similar power and coverage, GLAMRE has higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

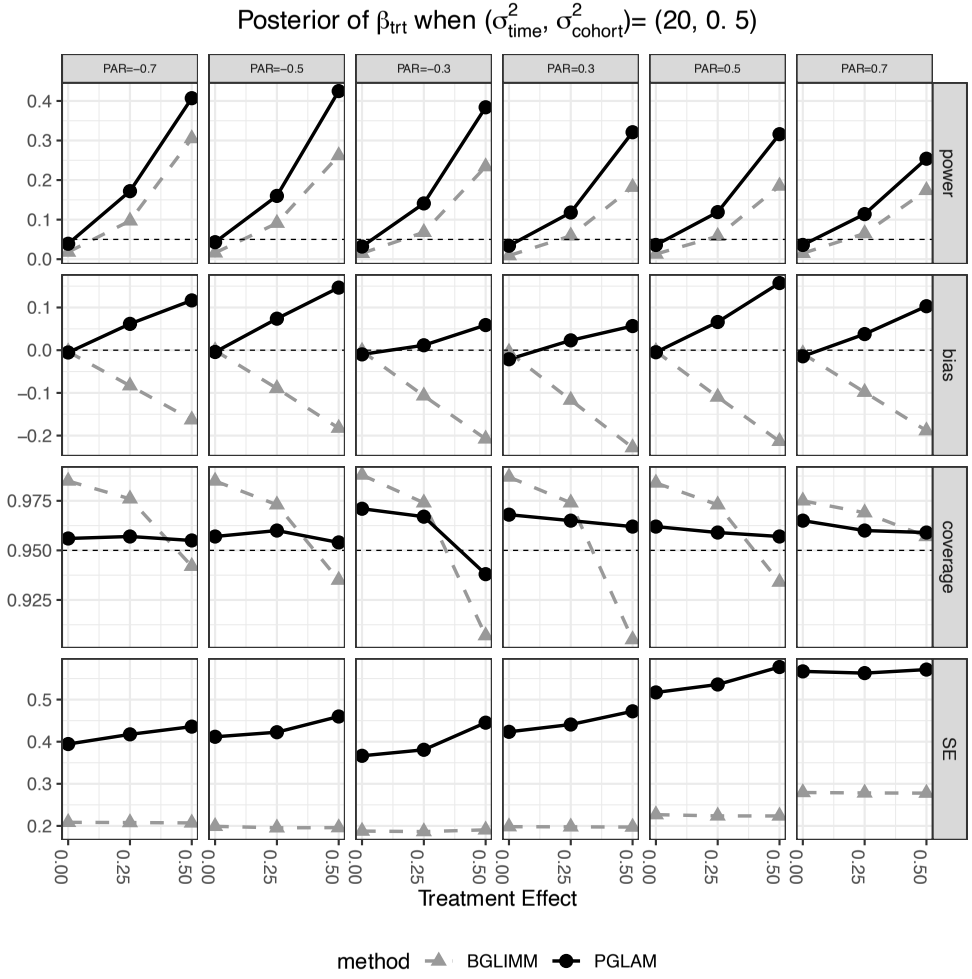


Figure B.3. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 0.5)$. Here GLAMRE has higher power, lower bias, better coverage but higher standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

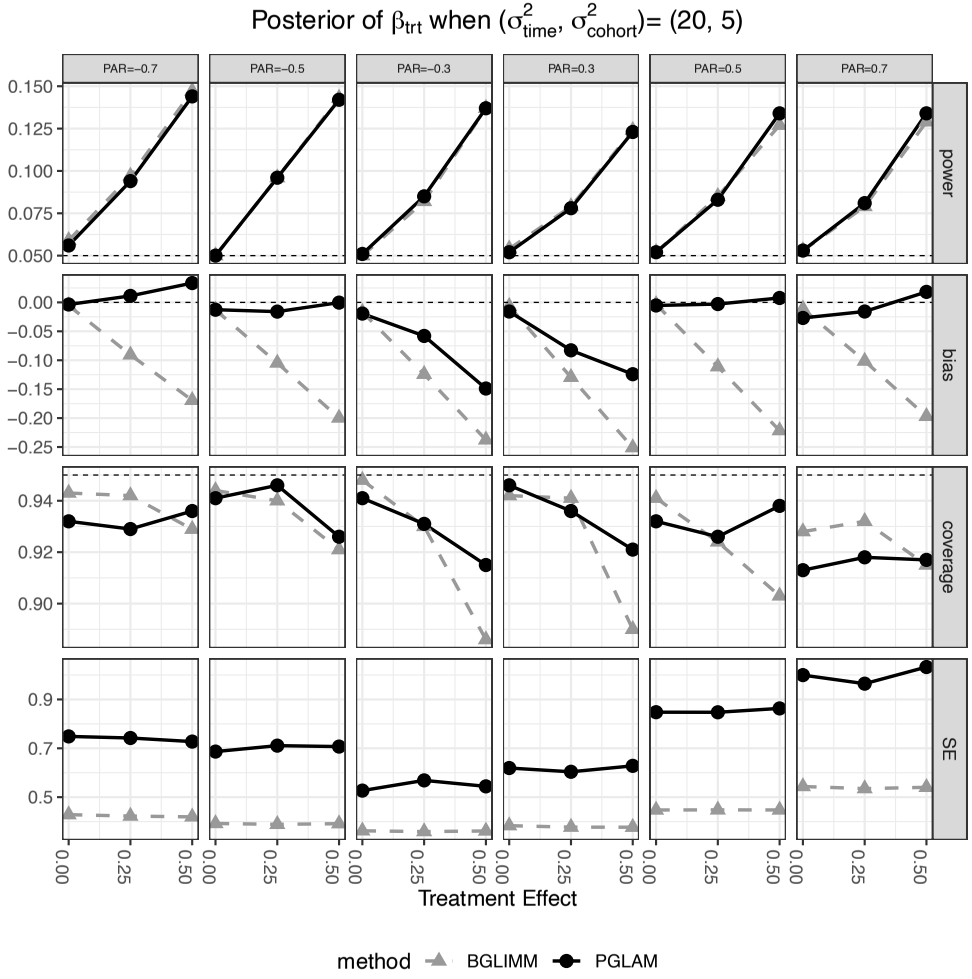


Figure B.4. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 5)$. Here GLAMRE has higher power, lower bias, better coverage but higher standard errors. Although both models have similar power and coverage, GLAMRE has higher bias and standard error. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

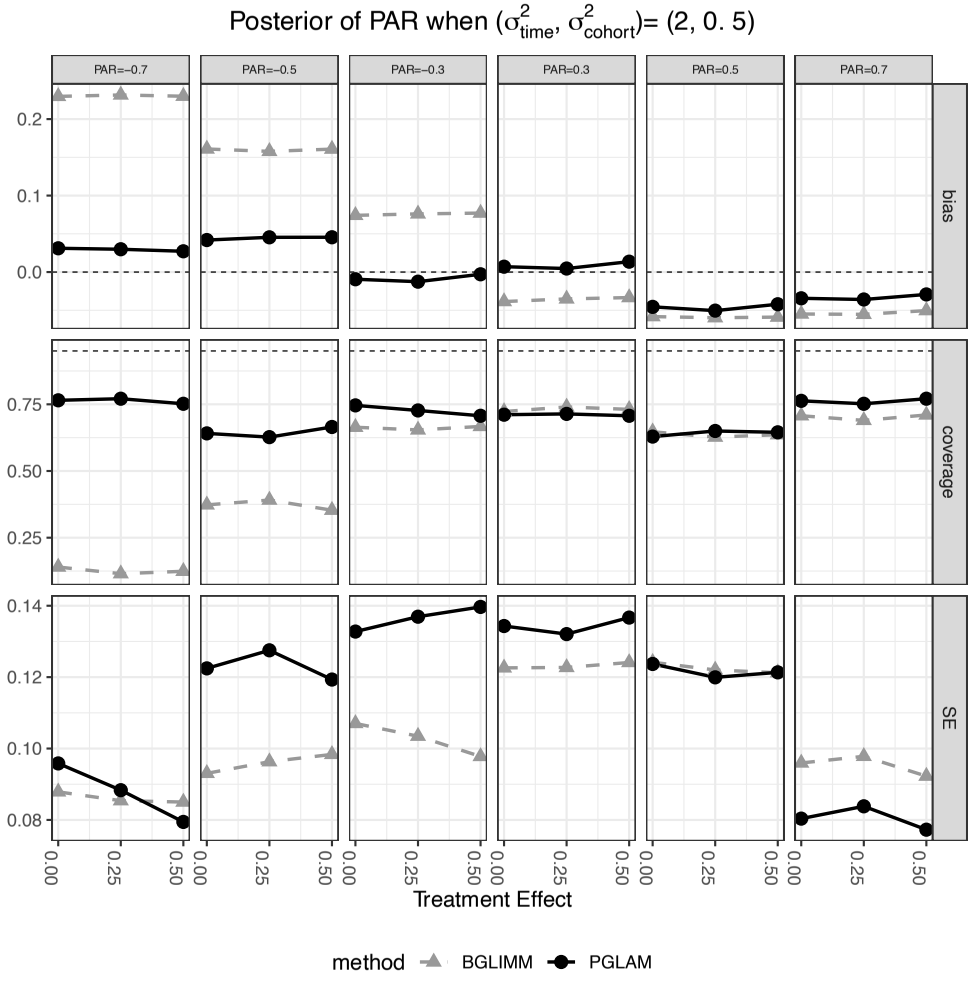


Figure B.5. Plots of bias, coverage and standard error of PAR estimates for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 0.5)$. Here GLAMRE has lower bias, better coverage and comparable standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

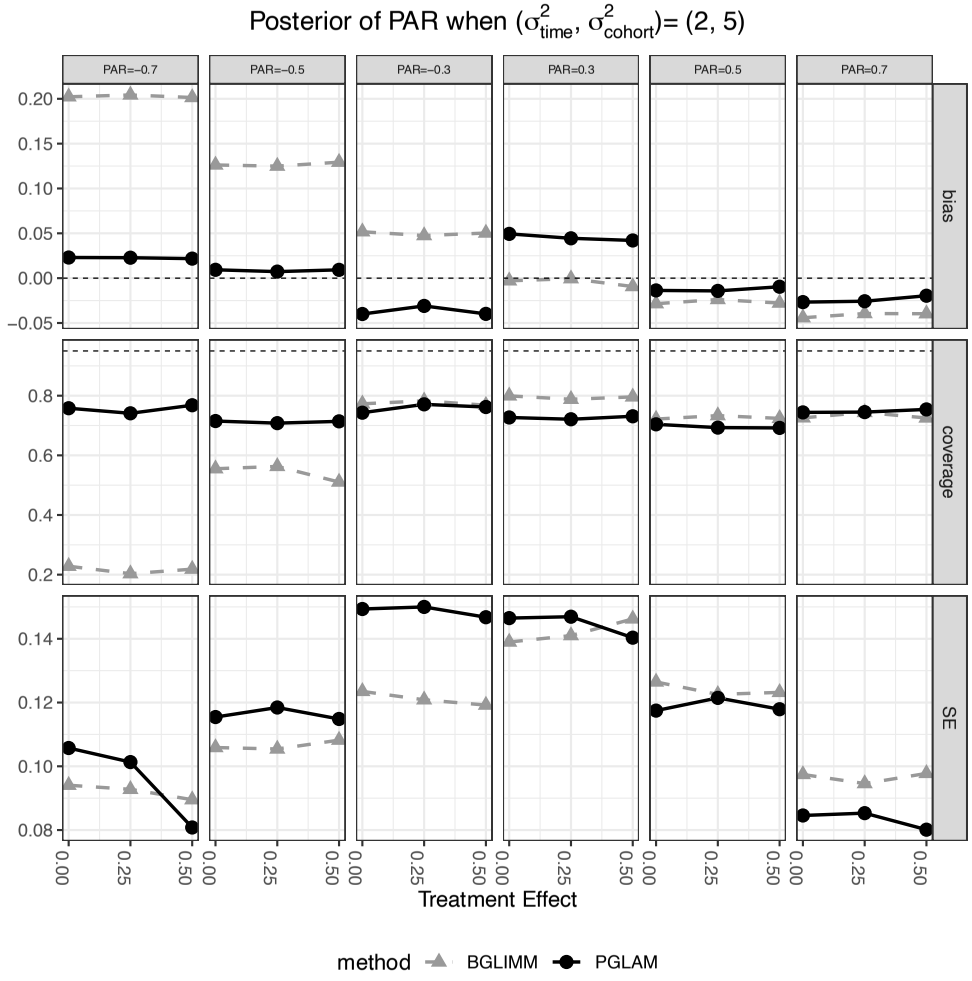


Figure B.6. Plots of bias, coverage and standard error of PAR estimates for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 5)$. Here GLAMRE has lower bias, better coverage and comparable standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

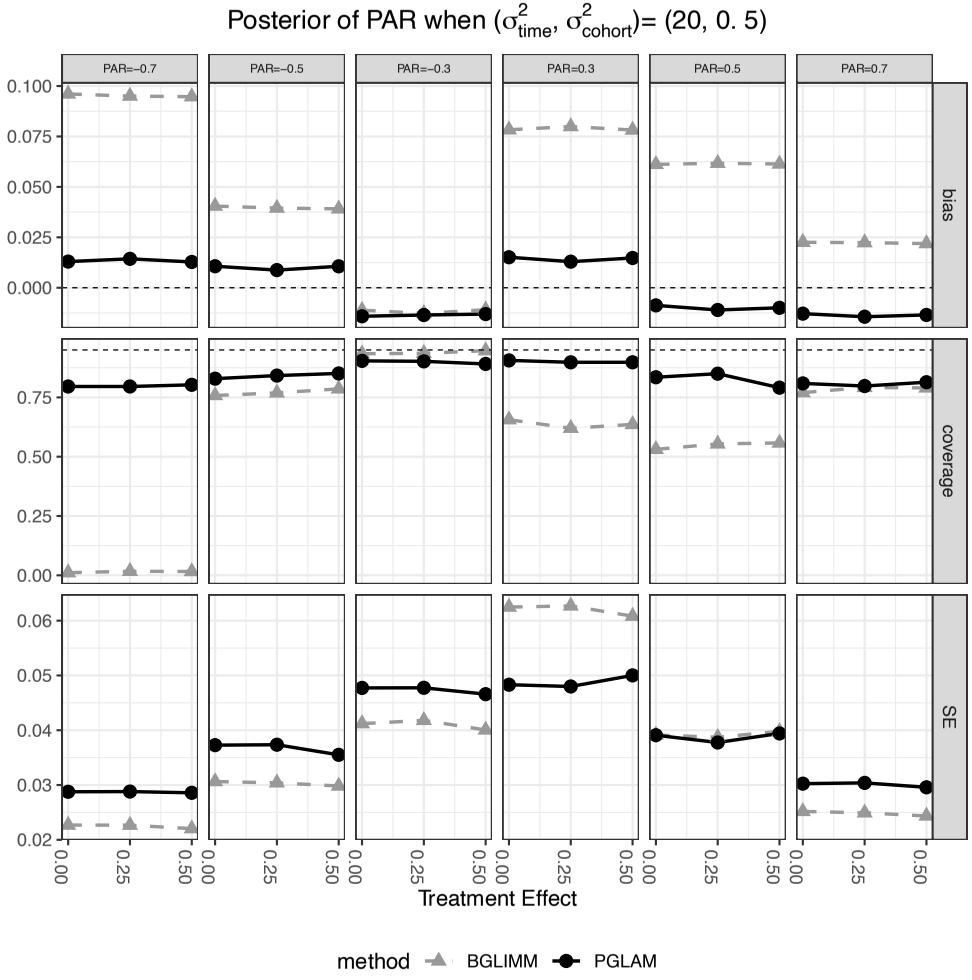


Figure B.7. Plots of bias, coverage and standard error of PAR estimates for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 0.5)$. Here GLAMRE has lower bias, better coverage and comparable standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

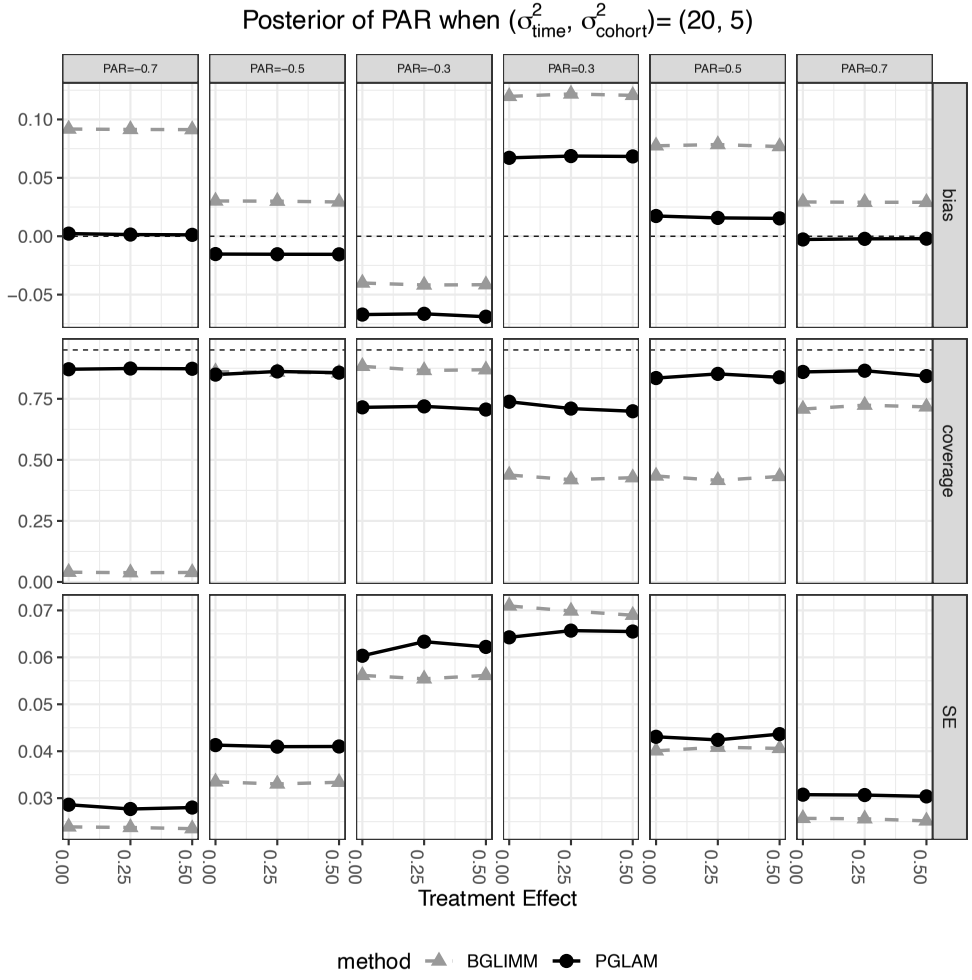


Figure B.8. Plots of bias, coverage and standard error of PAR estimates for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 5)$. Here GLAMRE has generally lower bias, better coverage and similar standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

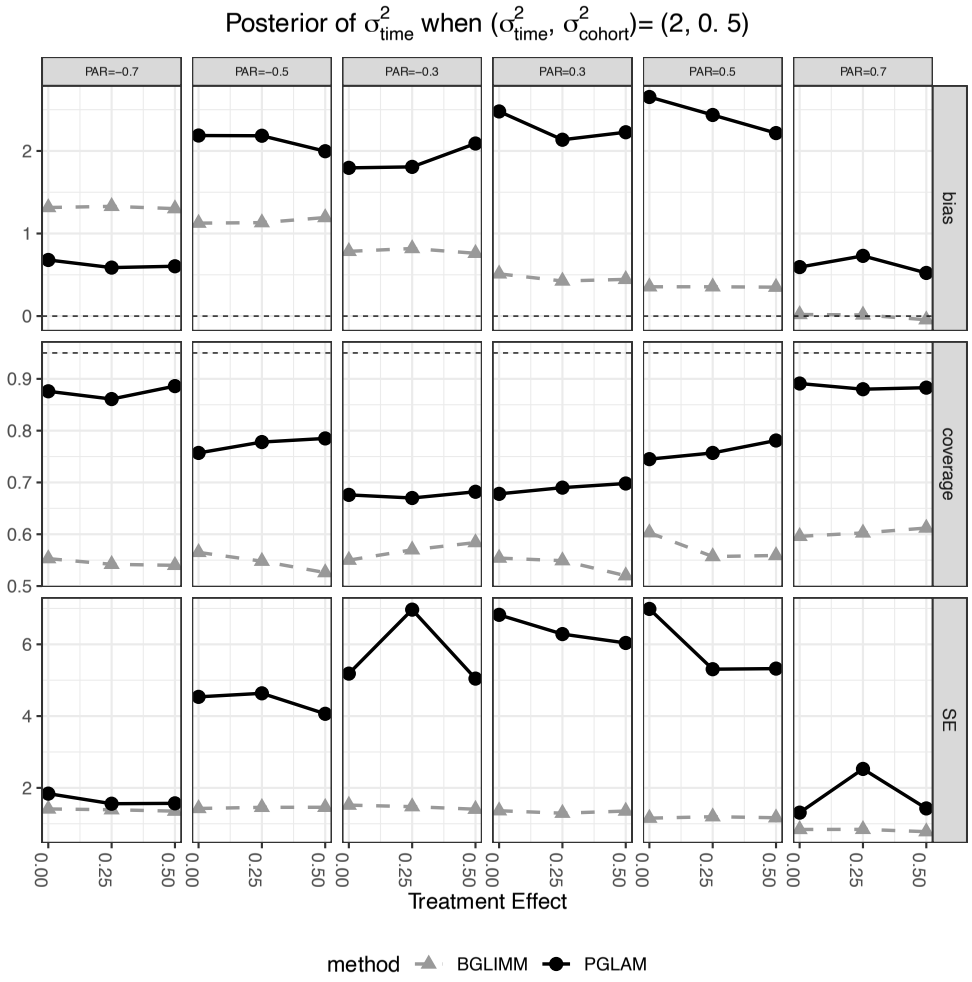


Figure B.9. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 0.5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

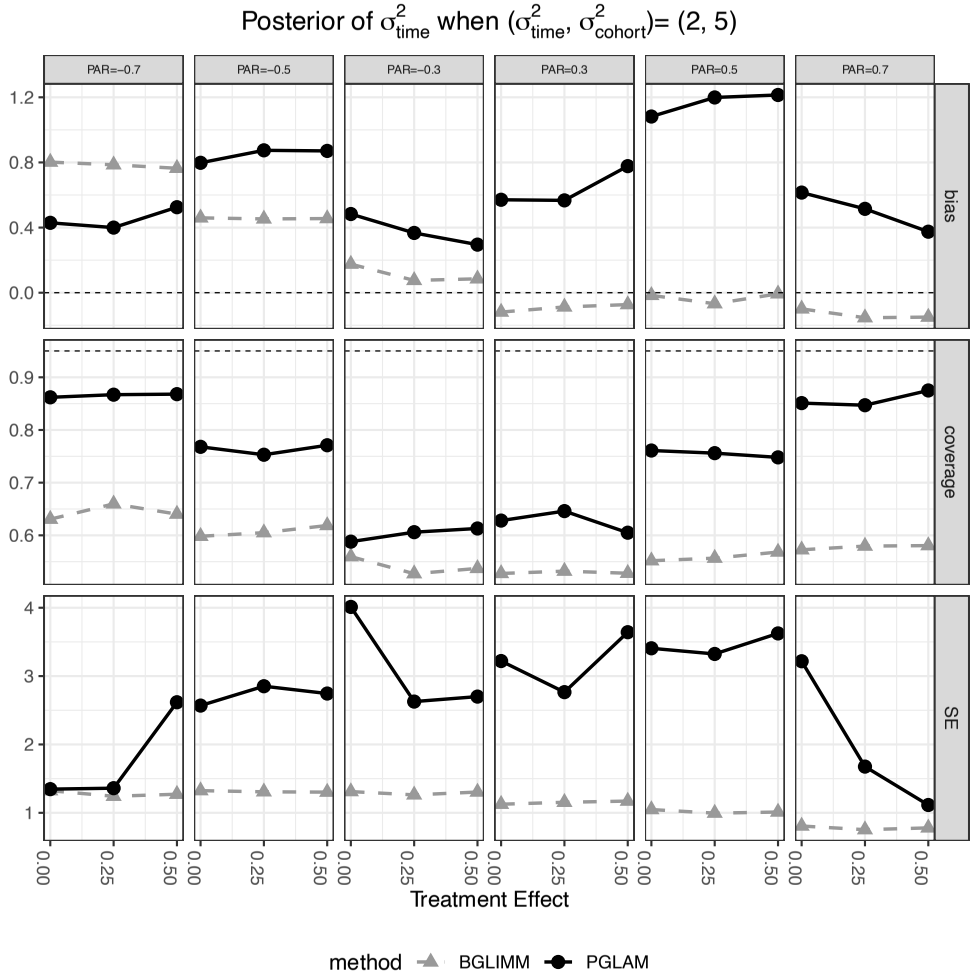


Figure B.10. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

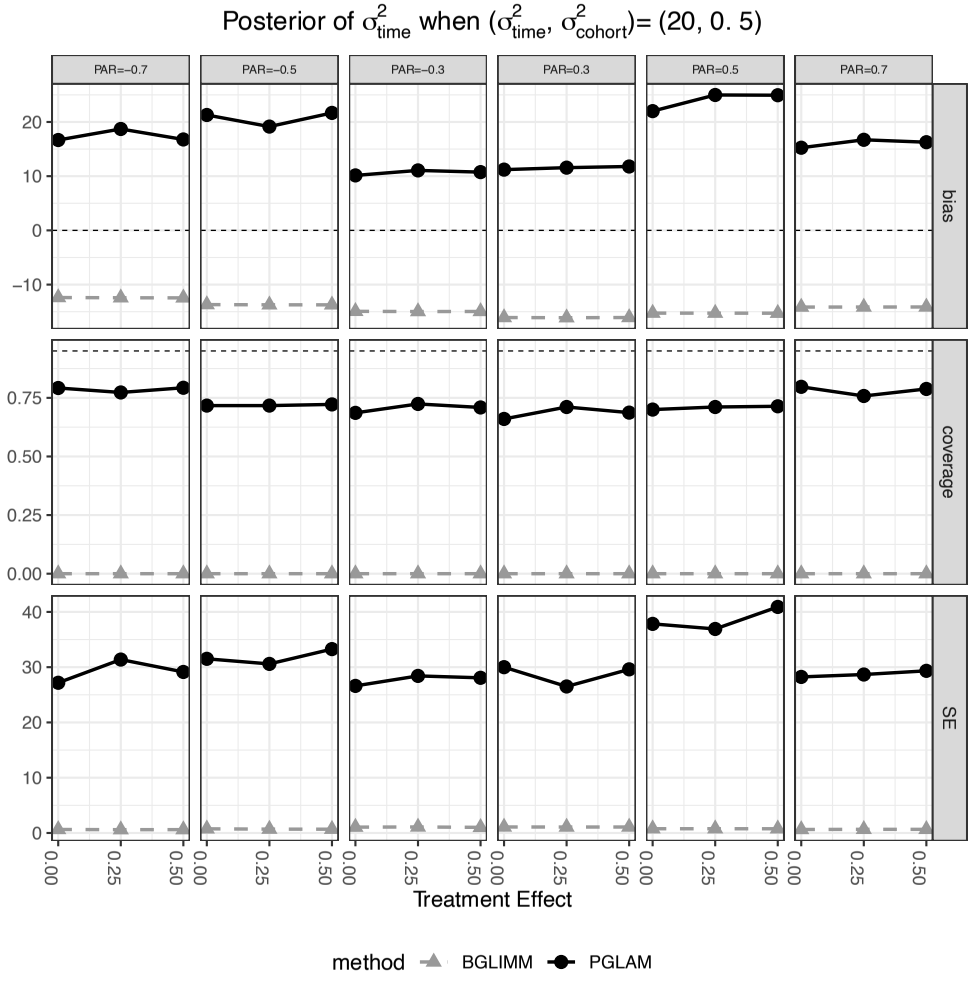


Figure B.11. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 0.5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

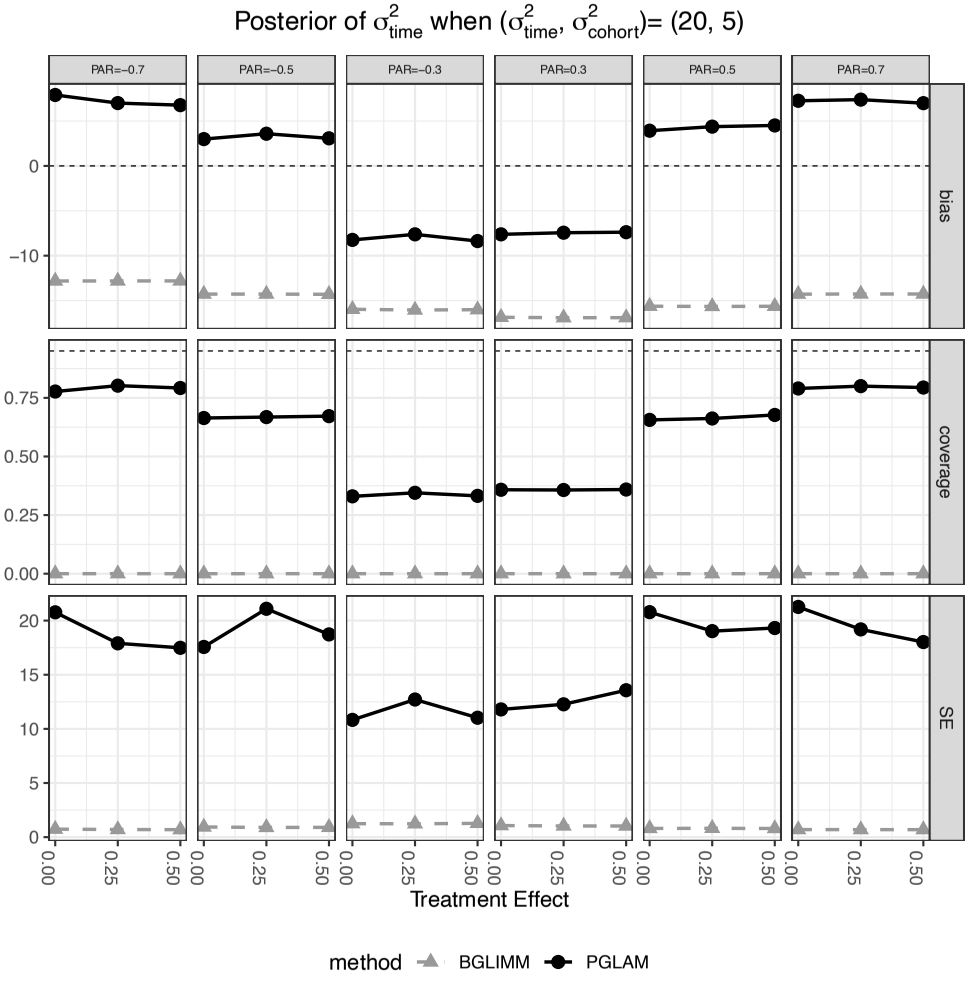


Figure B.12. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

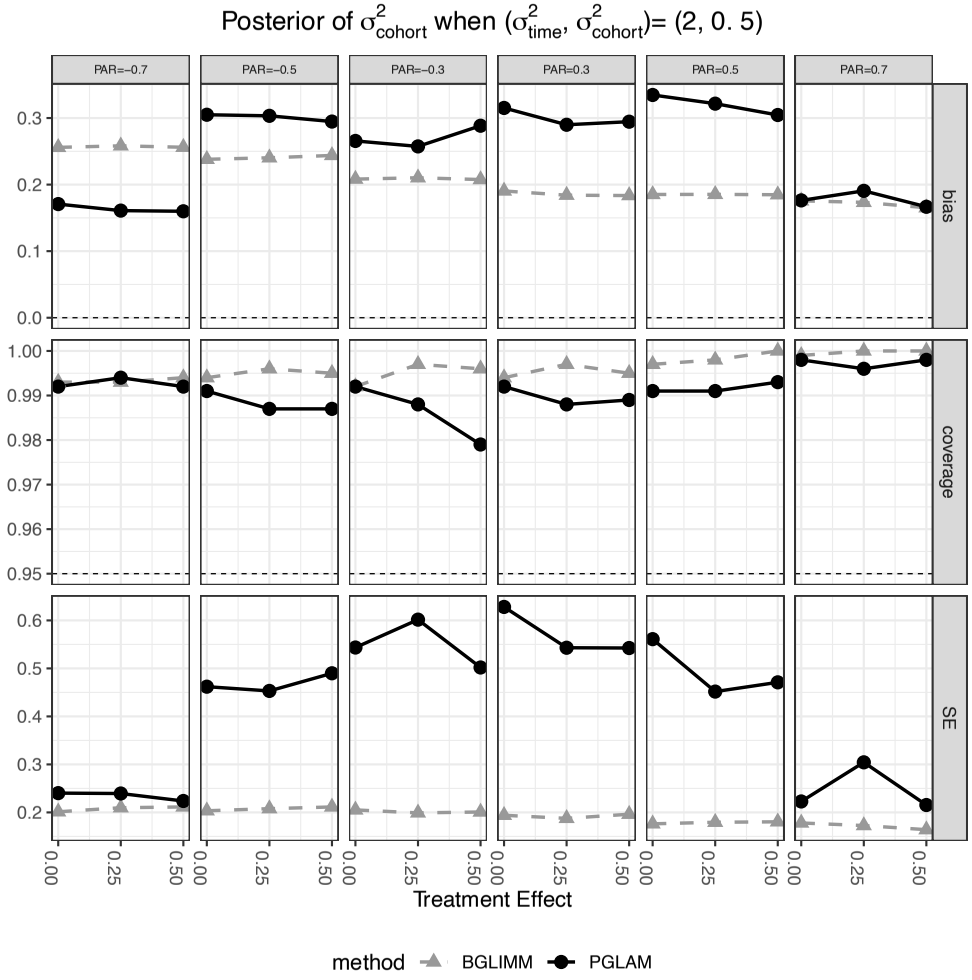


Figure B.13. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 0.5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

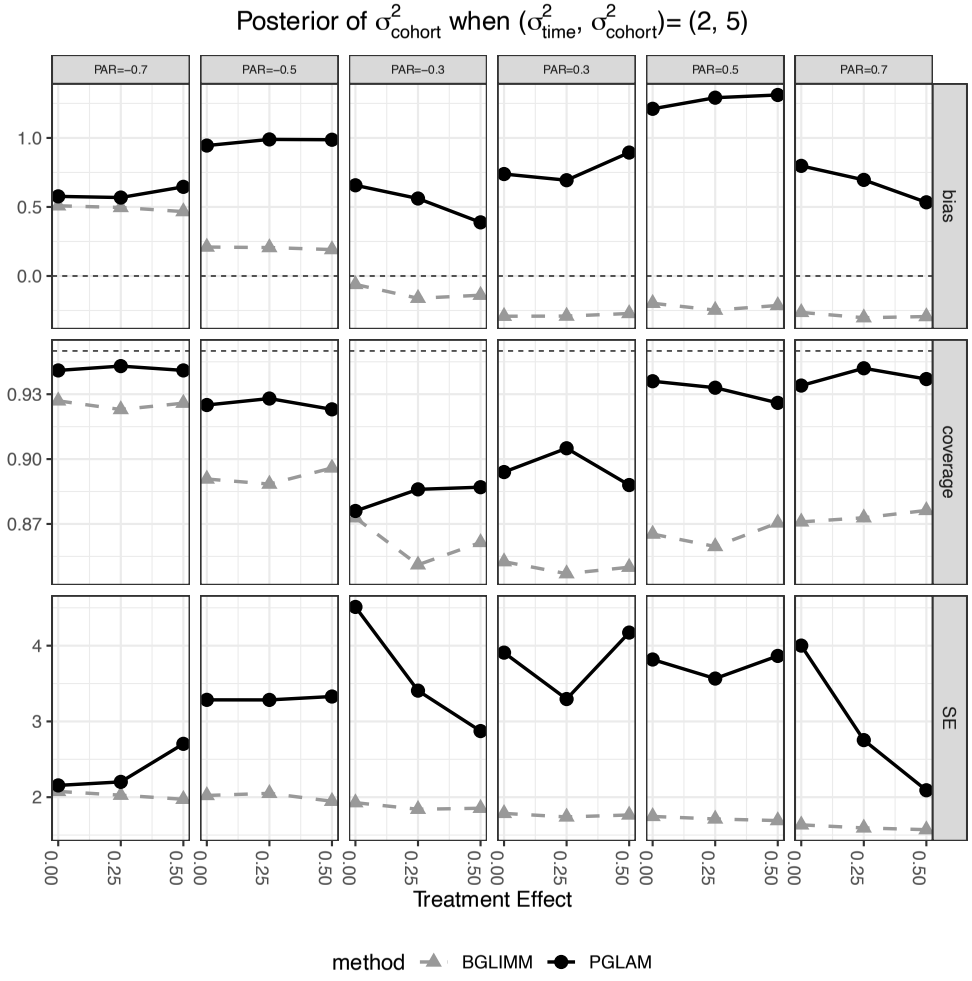


Figure B.14. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

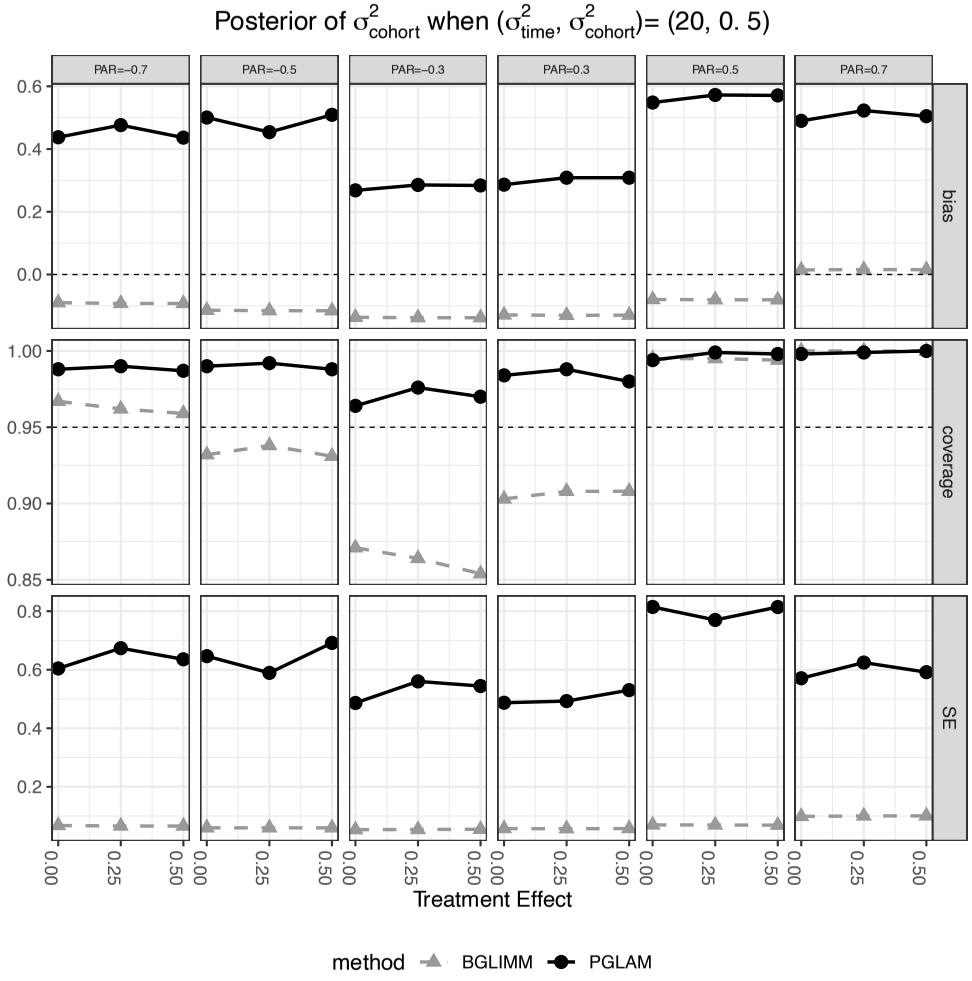


Figure B.15. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 0.5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

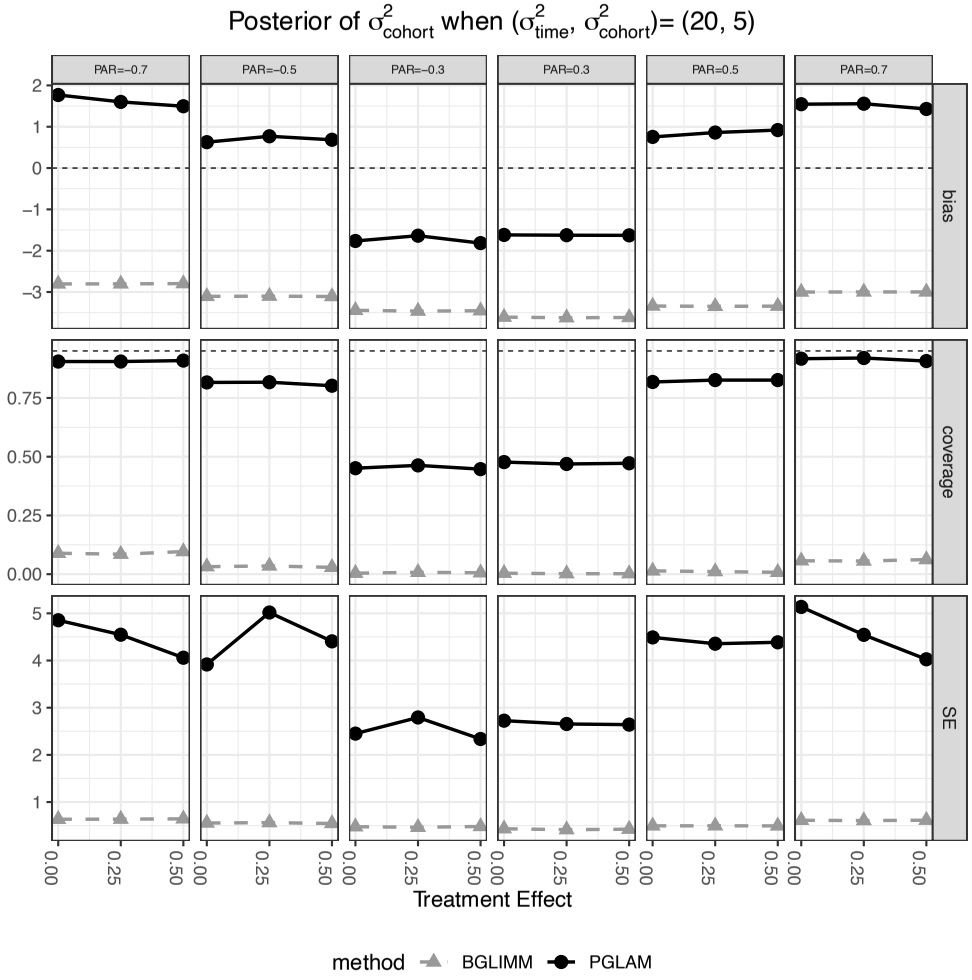


Figure B.16. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(1) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 5)$. GLAMRE has much better coverage, but at the expense of higher bias and standard errors. This is primarily due to better coverage in the other random effects, which translates to added variability in the estimation procedure.

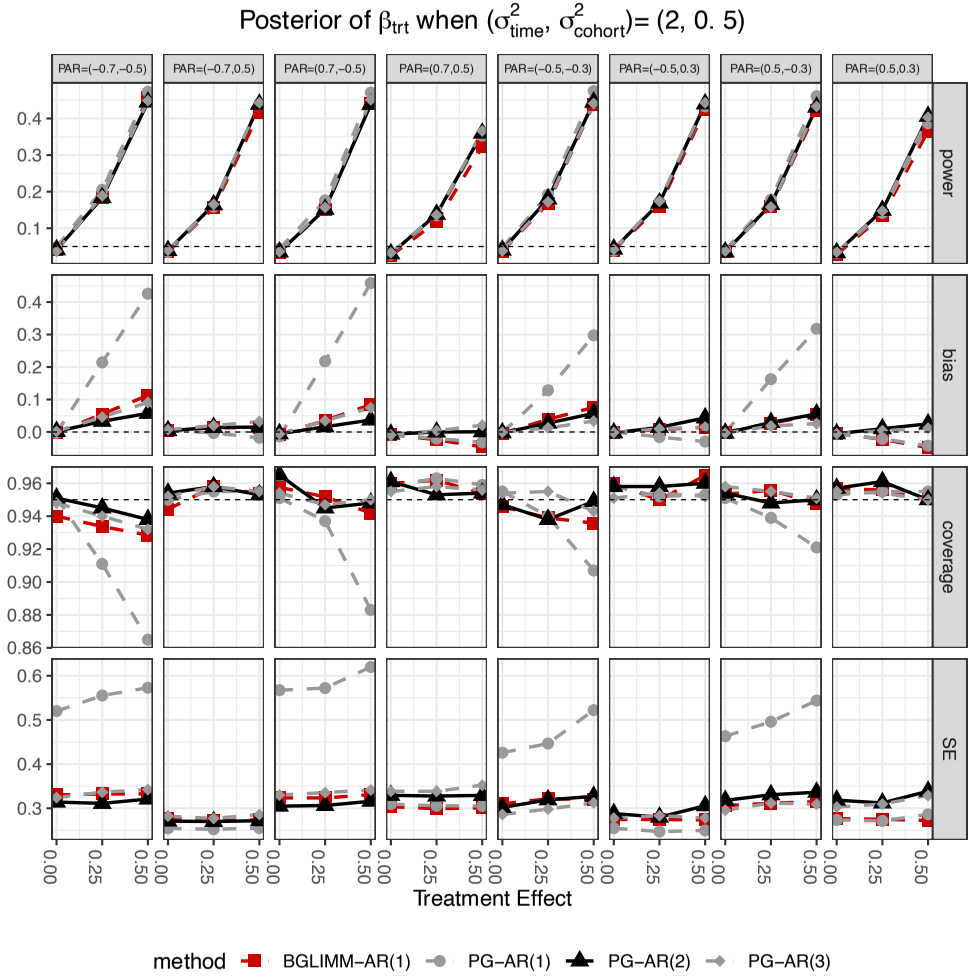


Figure B.17. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 0.5)$. Here all models but PGLMA-AR(1) have similar model performances. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

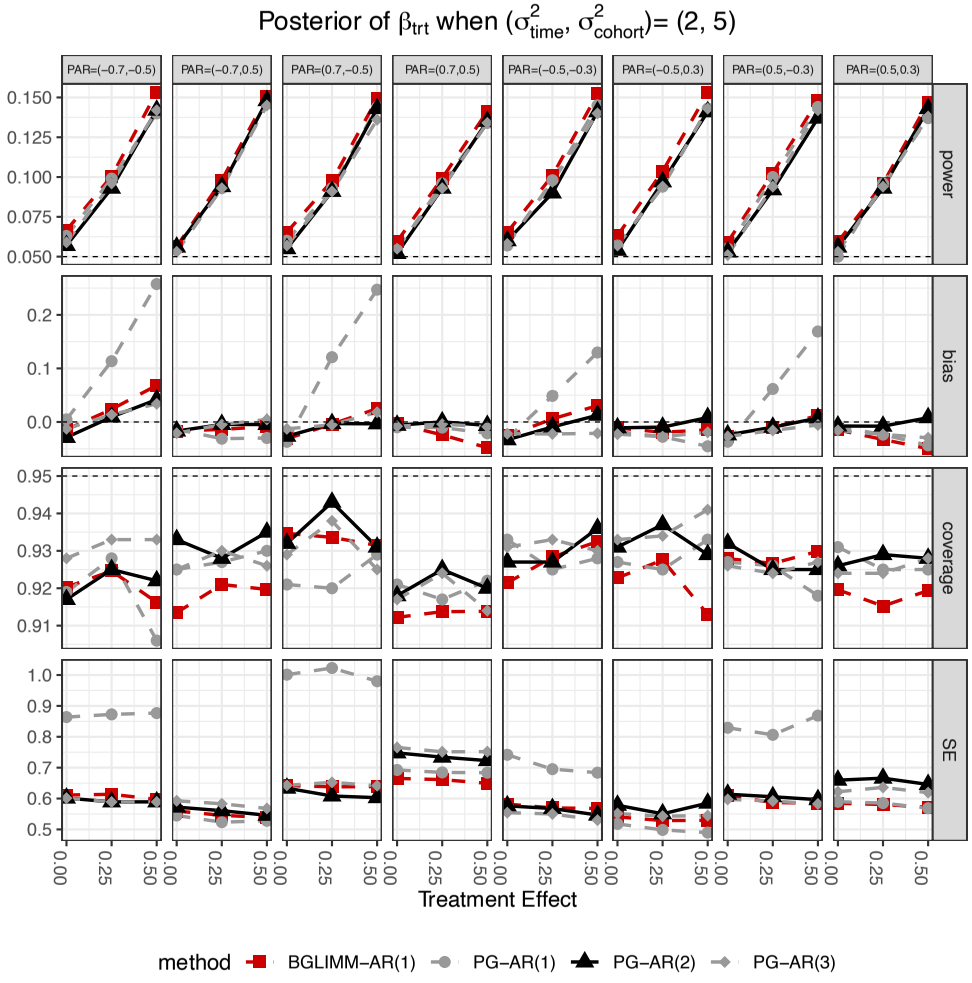


Figure B.18. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 5)$. Here all models but PGLMA-AR(1) have similar model performances. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

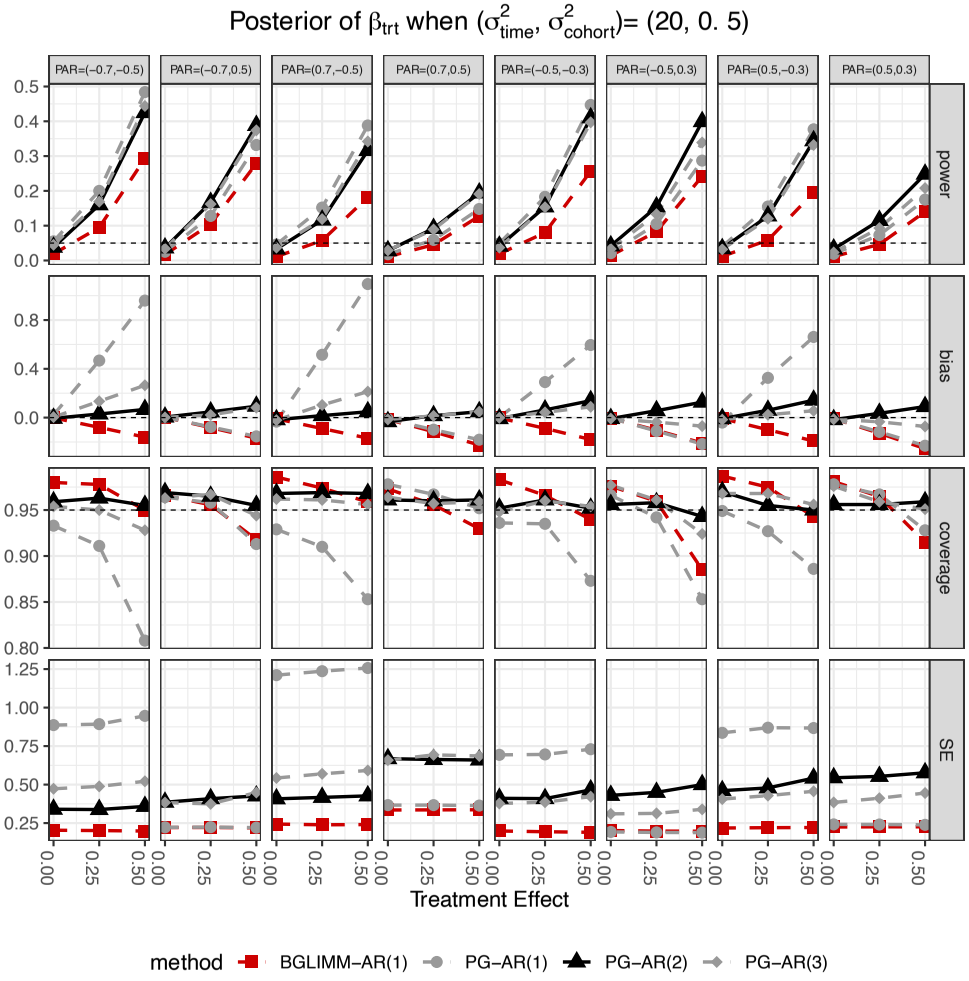


Figure B.19. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 0.5)$. Although GLAMRE-AR(2) has higher power, it also has higher standard errors. Here all models but PGLMA-AR(1) have similar bias and coverage performances. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

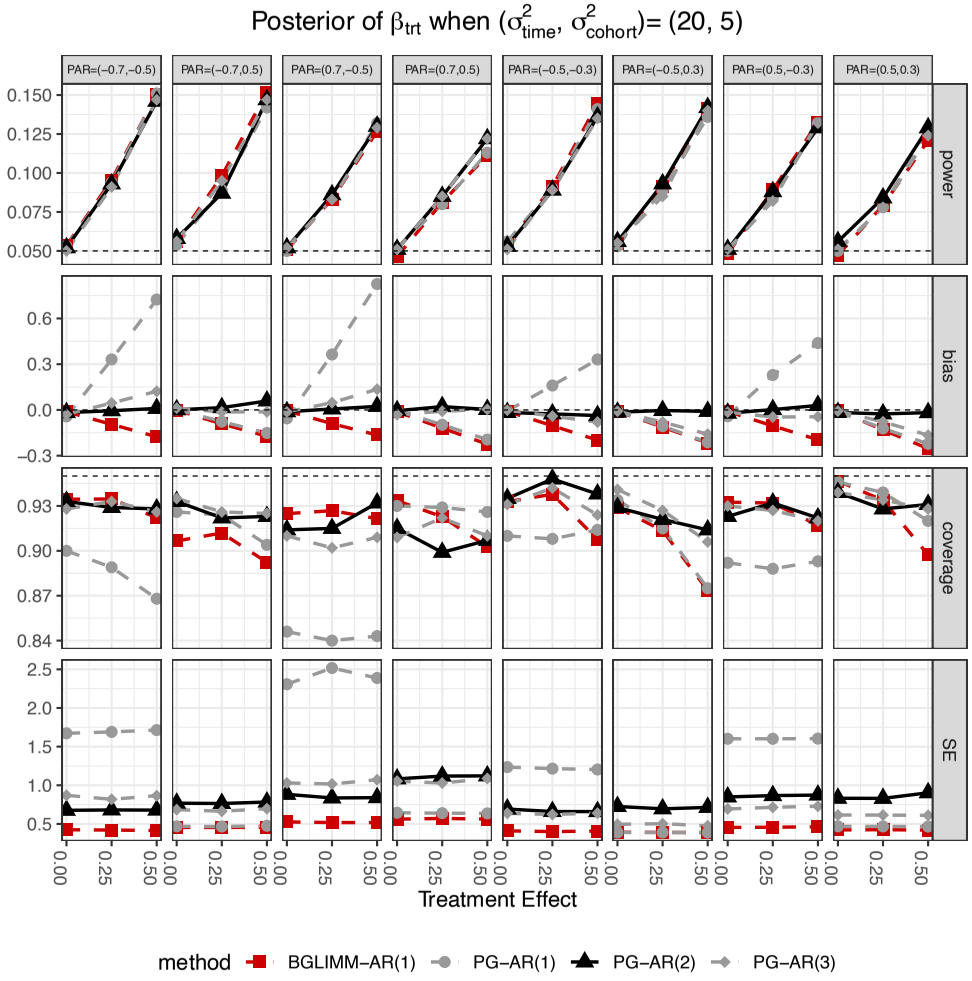


Figure B.20. Plots of power, bias, coverage and standard error of β_{trt} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 5)$. Here all models but PGLMA-AR(1) have similar model performances. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

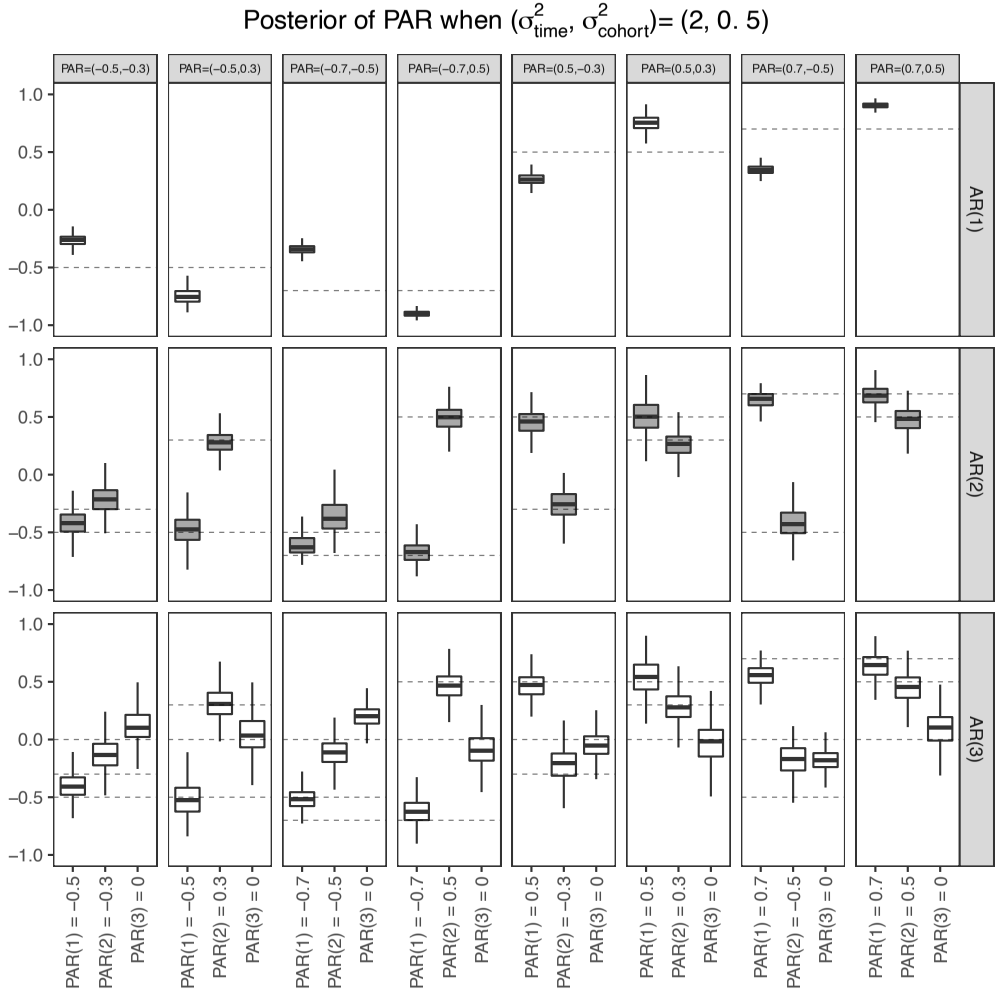


Figure B.21. Boxplots of the PARs for 1000 simulated datasets when $\beta_{\text{trt}} = 0, \sigma^2_{\text{time}} = 2, \sigma^2_{\text{cohort}} = 0.5$. The true correct model of AR(2) being the filled grey boxplots in the middle row. The horizontal dotted lines are the true PAR settings. The boxplots of the PARs in the AR(2) model (second row) cover the true values. Results are consistent across the other β_{trt} values.

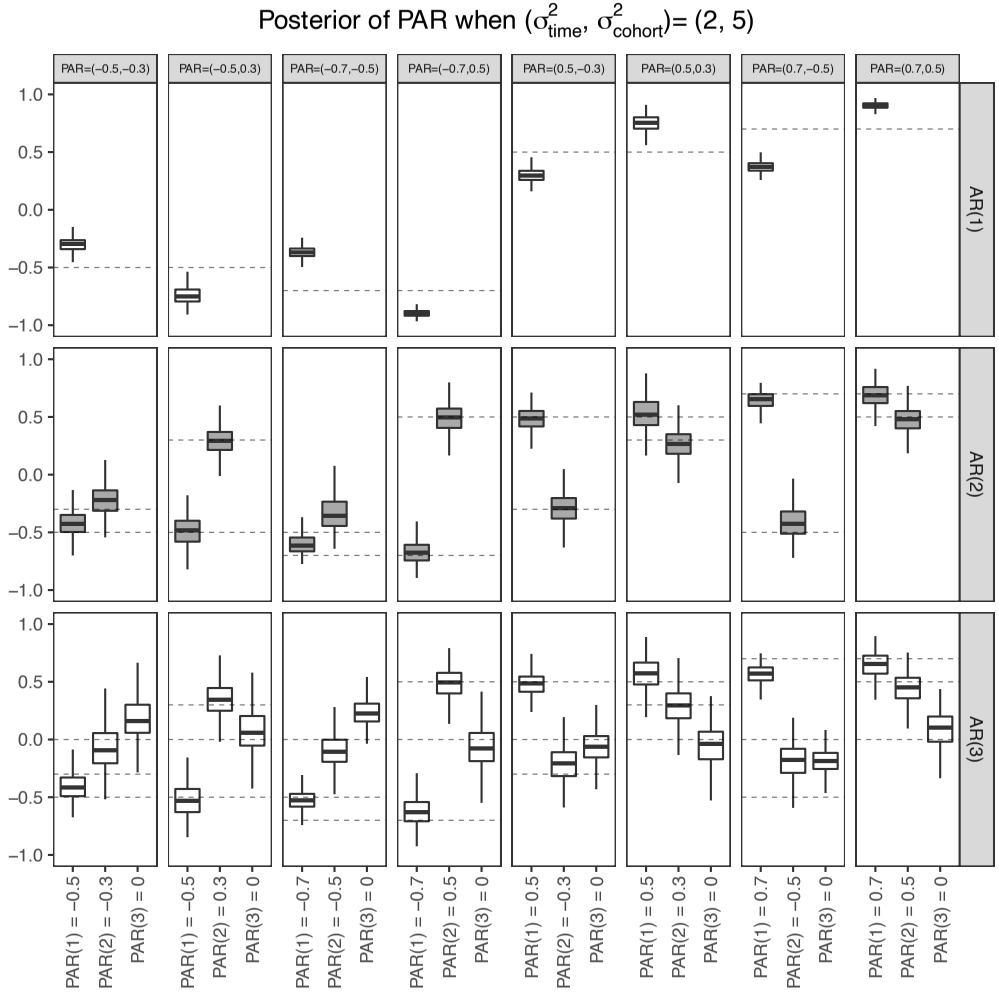


Figure B.22. Boxplots of the PARs for 1000 simulated datasets when $\beta_{\text{trt}} = 0, \sigma^2_{\text{time}} = 2, \sigma^2_{\text{cohort}} = 5$. The true correct model of AR(2) being the filled grey boxplots in the middle row. The horizontal dotted lines are the true PAR settings. The boxplots of the PARs in the AR(2) model (second row) cover the true values. Results are consistent across the other β_{trt} values.

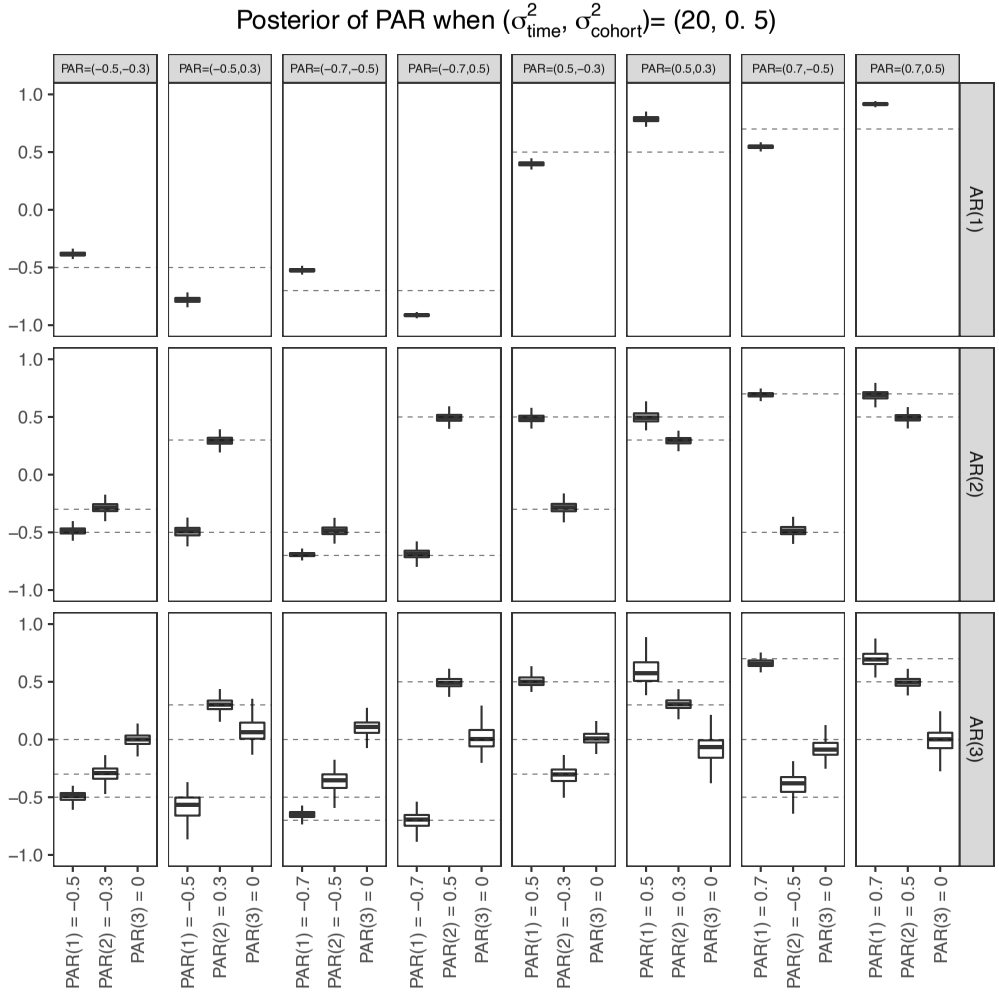


Figure B.23. Boxplots of the PARs for 1000 simulated datasets when $\beta_{\text{trt}} = 0, \sigma^2_{\text{time}} = 20, \sigma^2_{\text{cohort}} = 0.5$. The true correct model of AR(2) being the filled grey boxplots in the middle row. The horizontal dotted lines are the true PAR settings. The boxplots of the PARs in the AR(2) model (second row) cover the true values. Results are consistent across the other β_{trt} values.

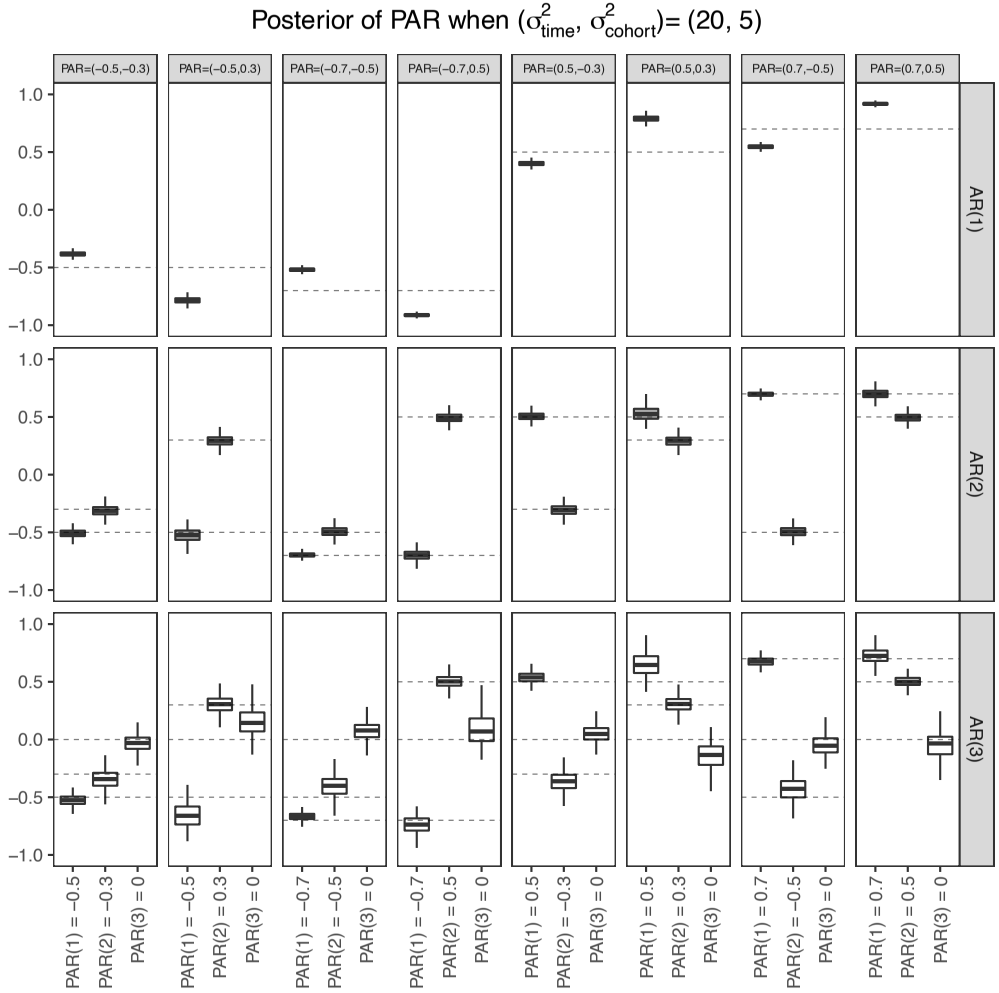


Figure B.24. Boxplots of the PARs for 1000 simulated datasets when $\beta_{\text{trt}} = 0, \sigma^2_{\text{time}} = 20, \sigma^2_{\text{cohort}} = 5$. The true correct model of AR(2) being the filled grey boxplots in the middle row. The horizontal dotted lines are the true PAR settings. The boxplots of the PARs in the AR(2) model (second row) cover the true values. Results are consistent across the other β_{trt} values.

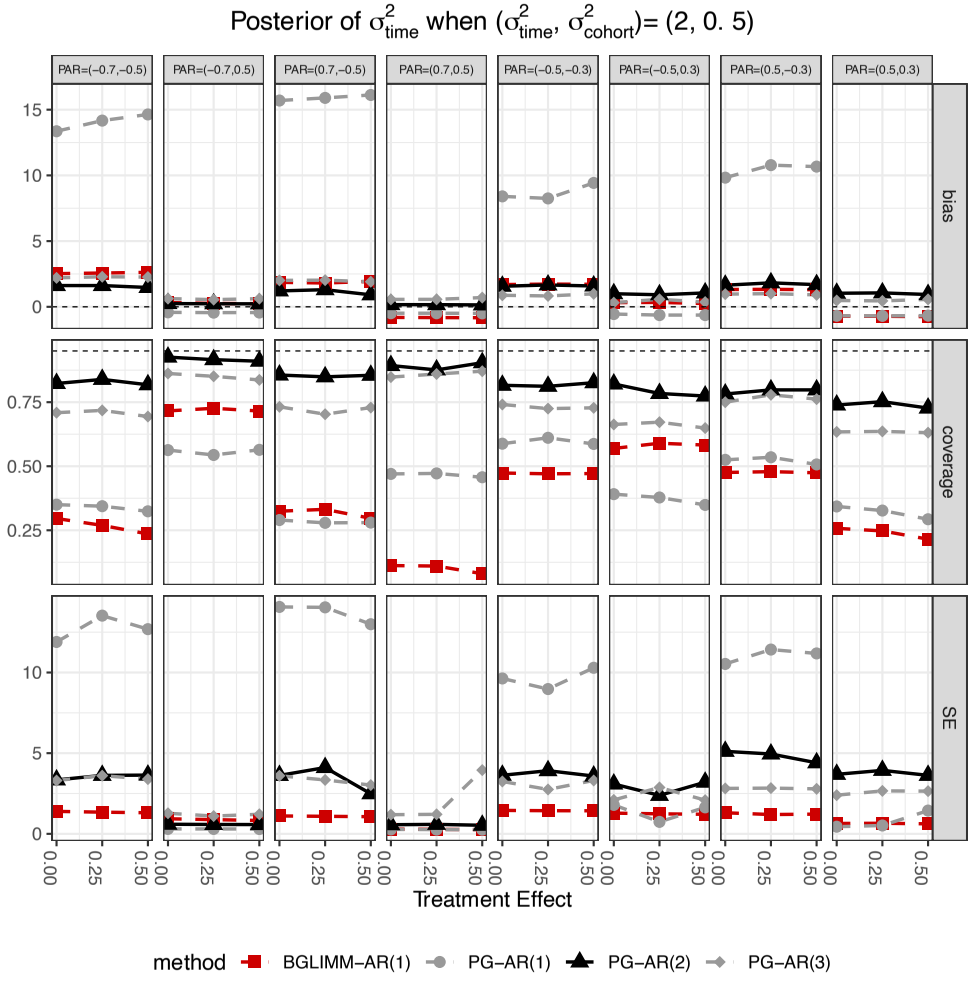


Figure B.25. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 0.5)$. GLAMRE-AR(2) has better coverage with low bias and standard errors. On the other hand, BGLIMM has consistent undercoverage. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

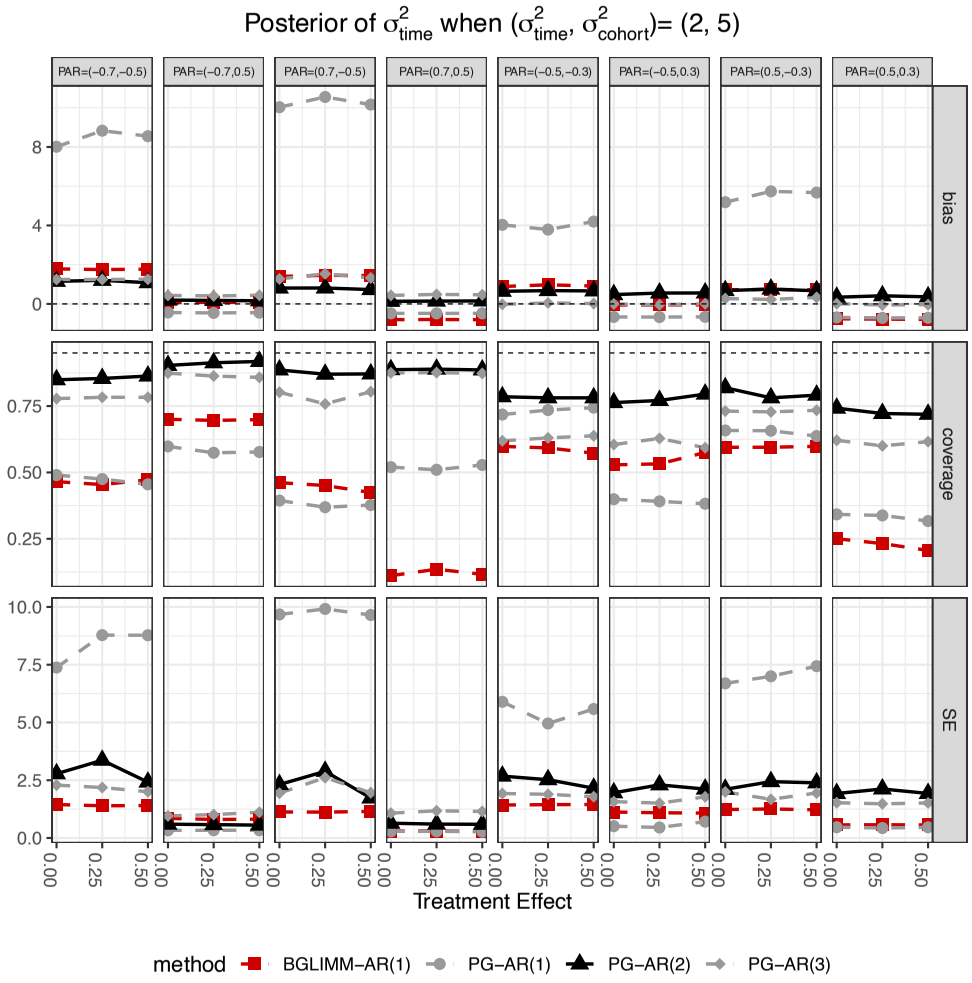


Figure B.26. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 5)$. GLAMRE-AR(2) has better coverage with low bias and standard errors. On the other hand, BGLIMM has consistent undercoverage. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

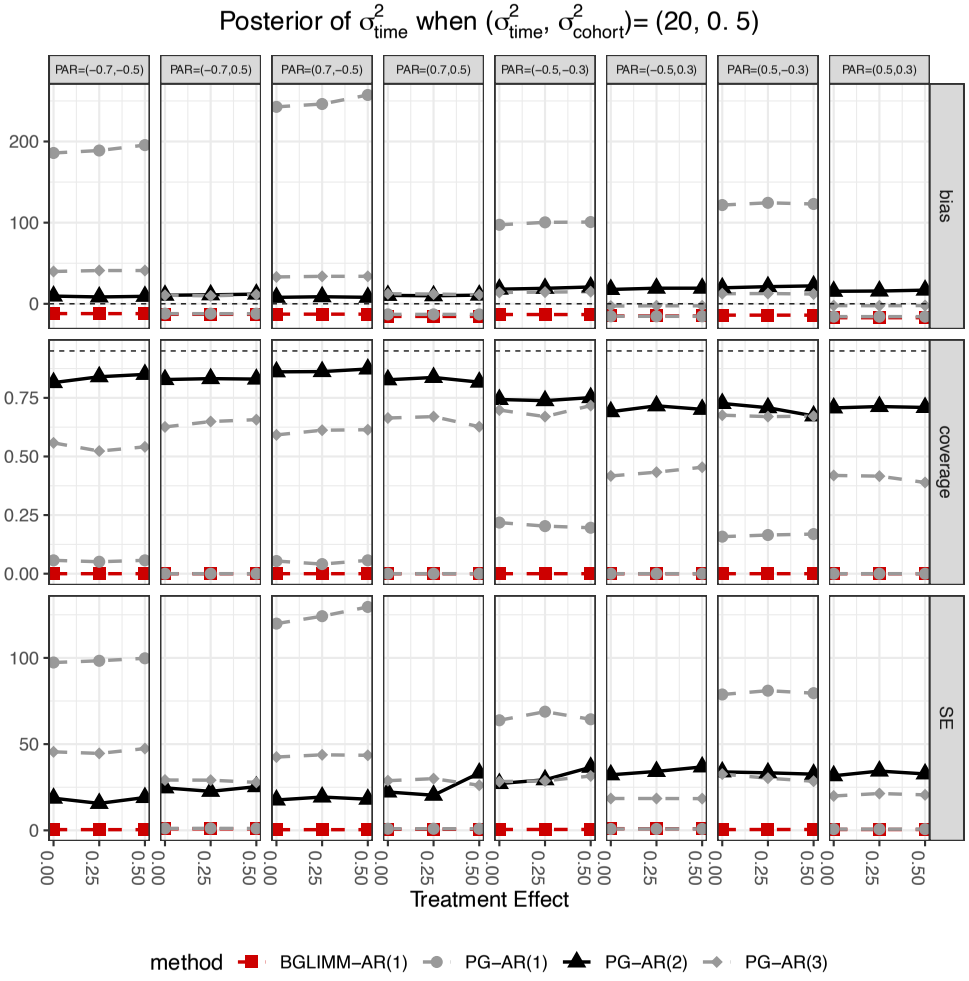


Figure B.27. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 0.5)$. GLAMRE-AR(2) has better coverage with low bias and standard errors. On the other hand, BGLIMM has consistent undercoverage. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

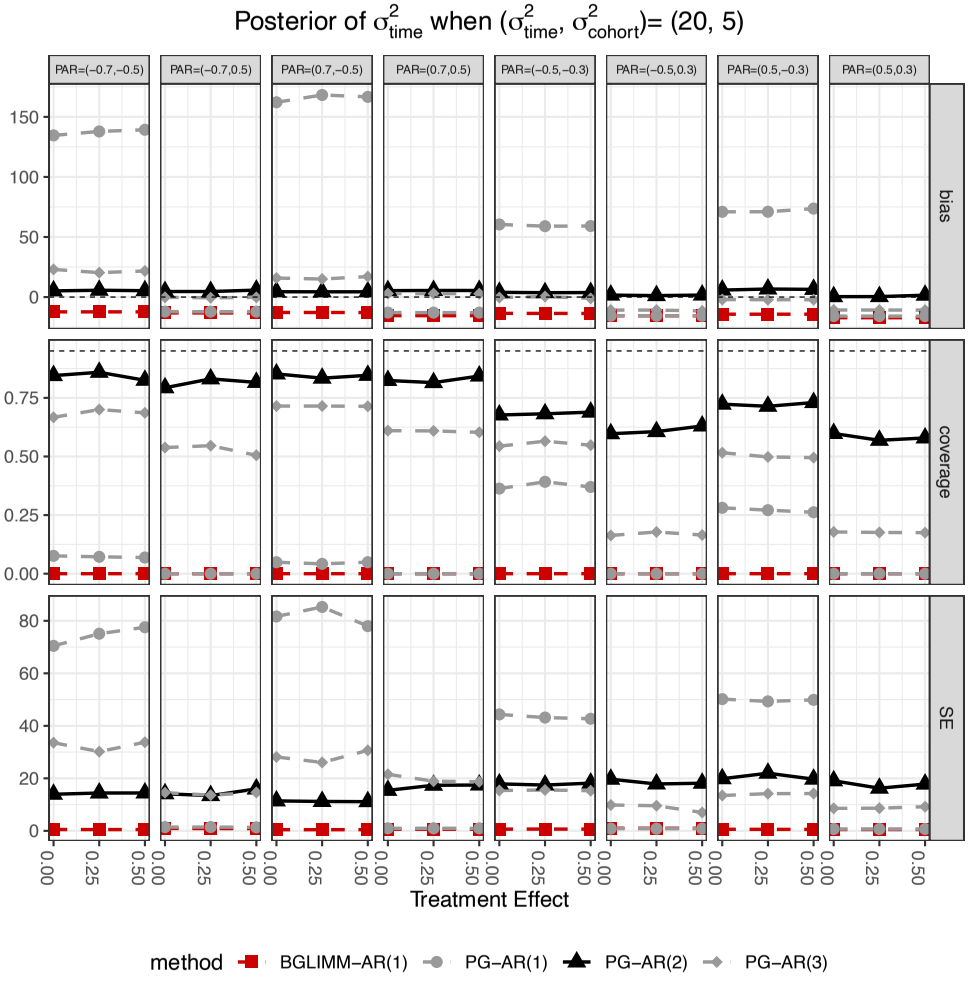


Figure B.28. Plots of bias, coverage and standard error of σ^2_{time} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 5)$. GLAMRE-AR(2) has better coverage with low bias and standard errors. On the other hand, BGLIMM has consistent undercoverage. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

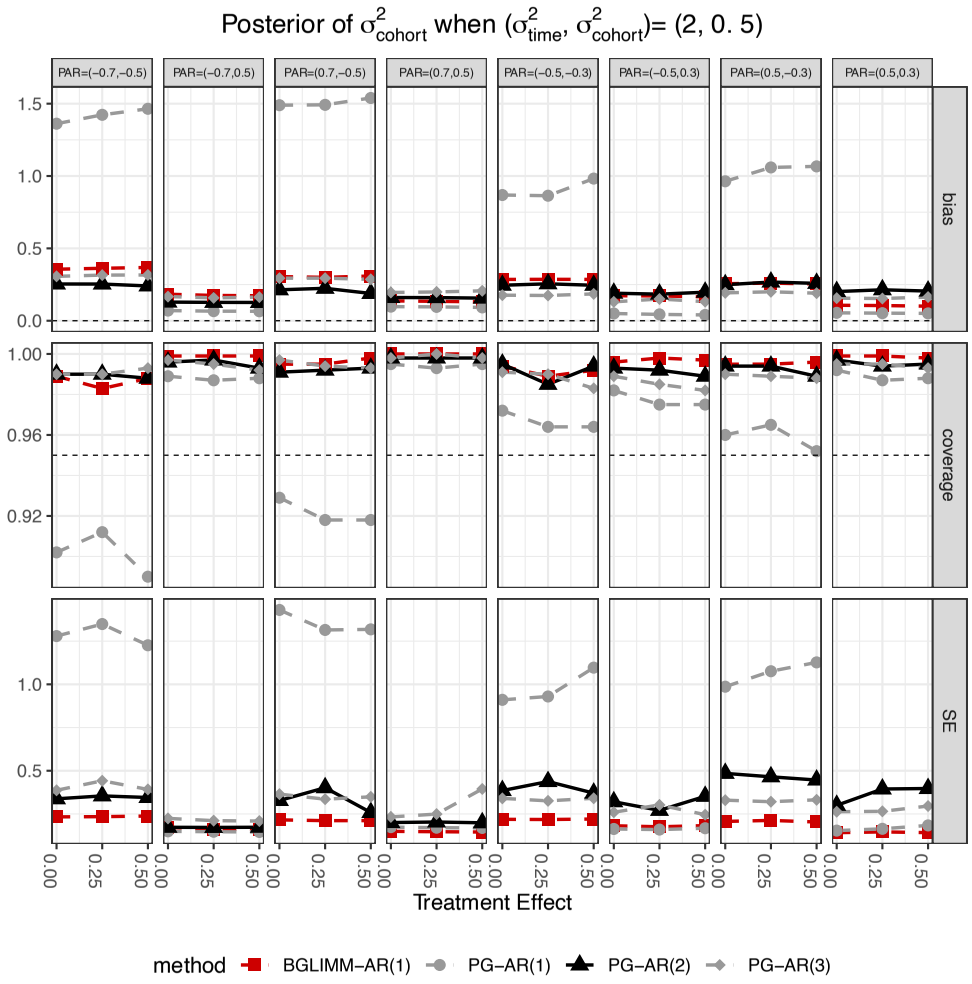


Figure B.29. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 0.5)$. GLAMRE-AR(2) has better coverage with low bias and standard errors. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

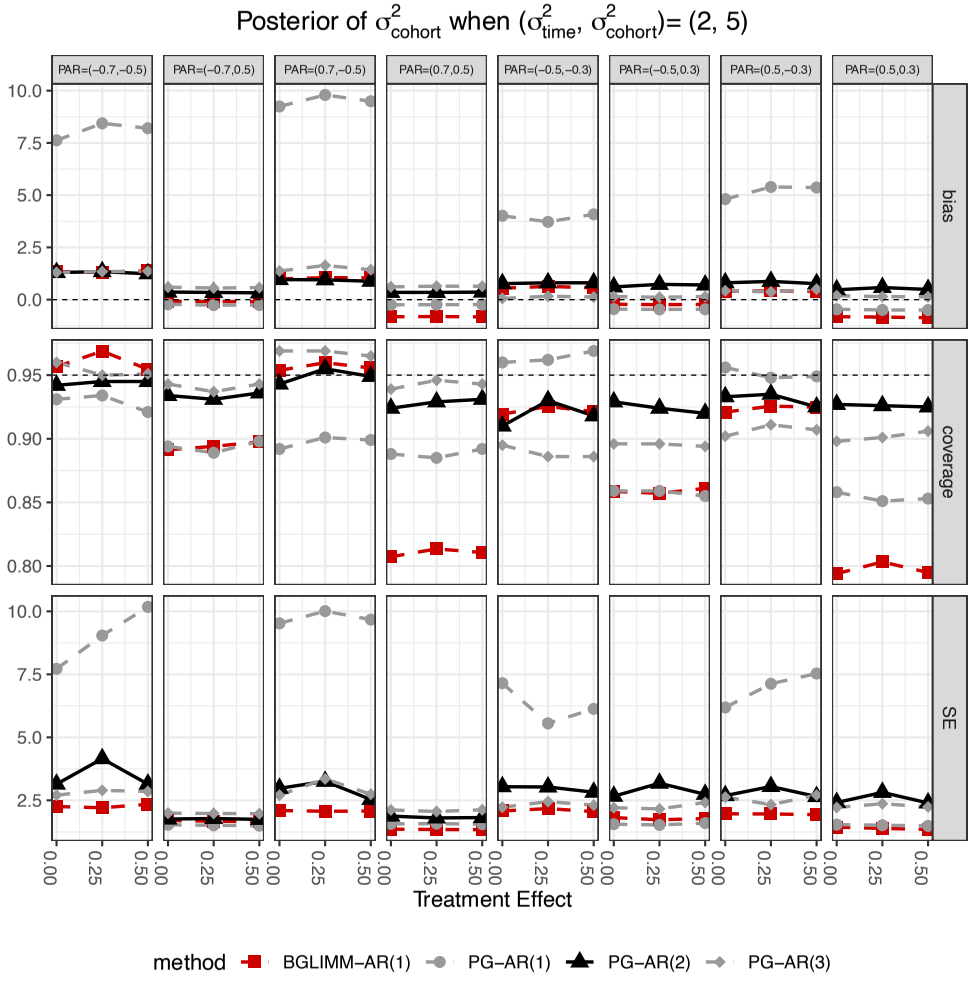


Figure B.30. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (2, 5)$. Although all models have similar biases, GLAMRE-AR(2) generally has better coverage while maintaining comparable standard errors. On the other hand, BGLIMM does suffer from undercoverage. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

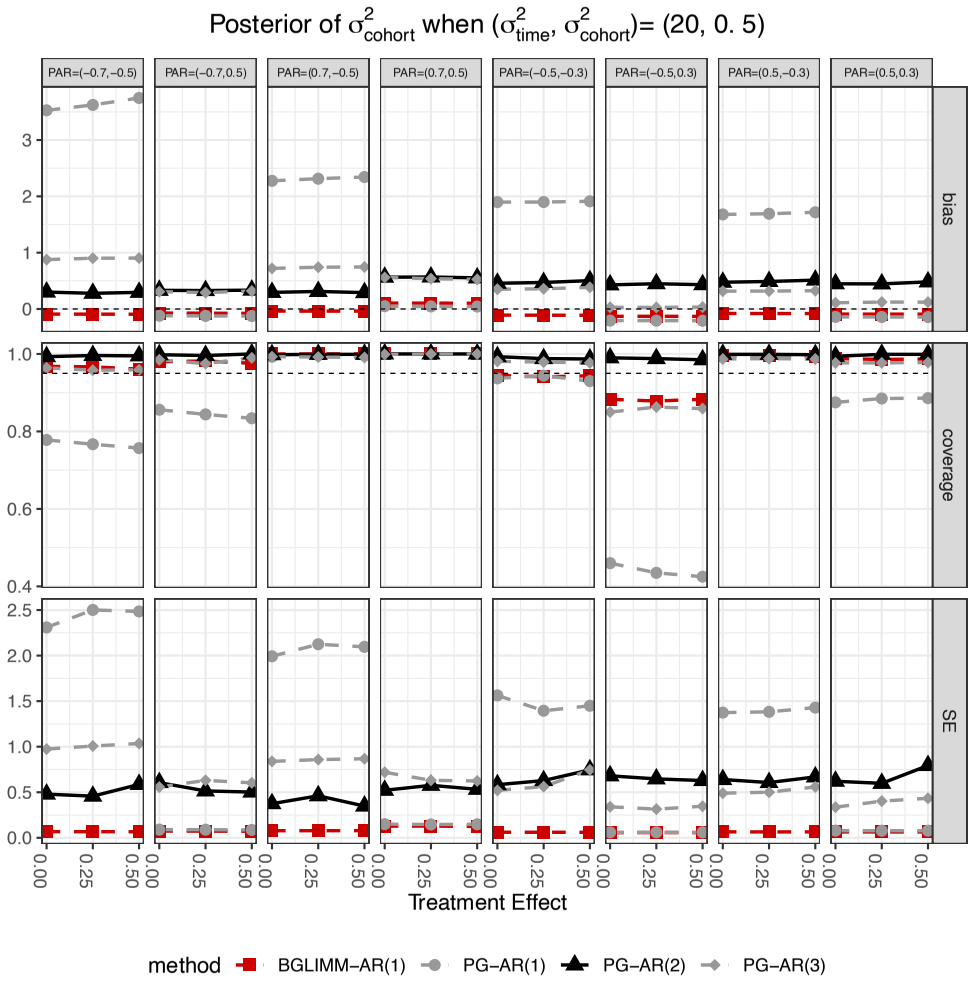


Figure B.31. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 0.5)$. All models but GLAMRE-AR(1) have similar biases and coverage. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

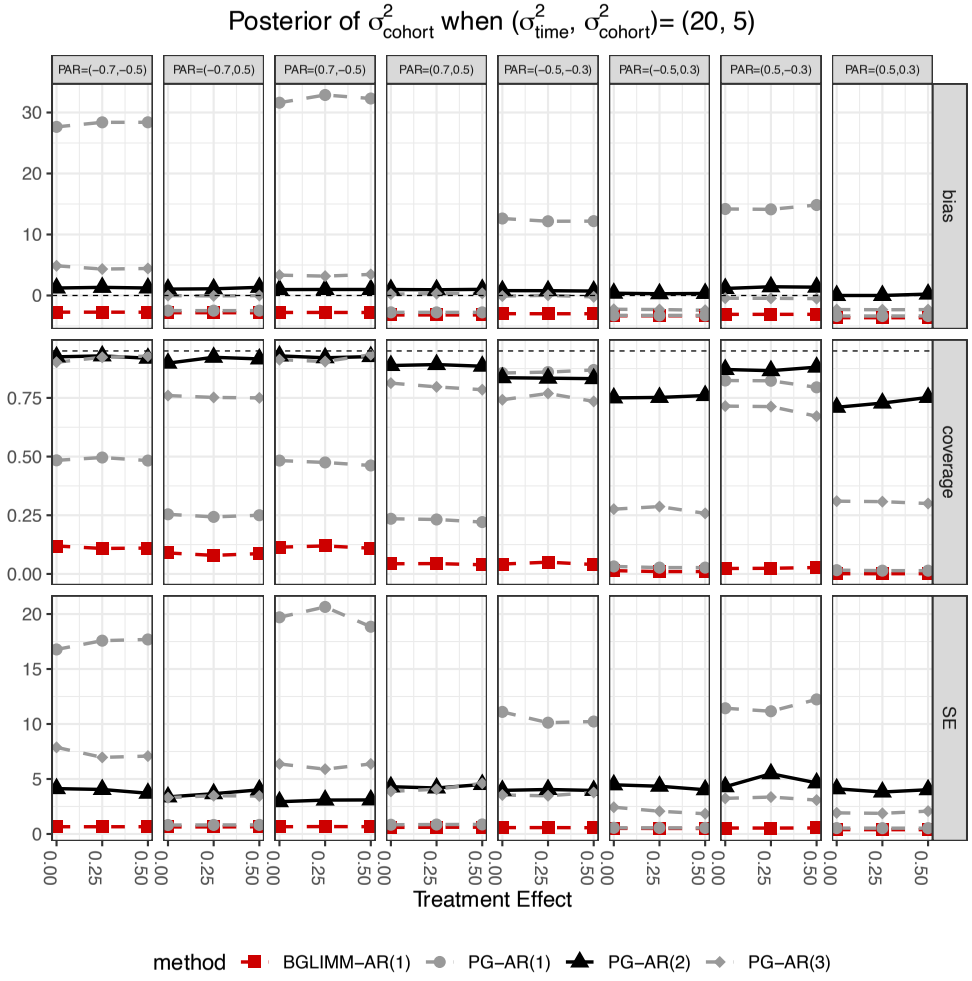


Figure B.32. Plots of bias, coverage and standard error of σ^2_{cohort} for 1000 simulated AR(2) datasets, when $(\sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}) = (20, 5)$.GLAMRE-AR(2) has lowest bias and best coverage, at the expense of slightly elevated standard errors. Due to model misspecification, GLAMRE-AR(1) has higher bias and lower coverage with higher standard errors.

Table B.1. Posterior summaries of PARs under different partitions with varying AR(k) settings as well as σ^2_{time} and σ^2_{cohort}.The 95% CIs of the PAR(3)'s for AR(3) cover 0. Further, the DIC for AR(2) for both the weekday and weekend partition is the lowest, compelling us to model the weekday and weekend correlation structure with an AR(2).

Method	AR(k)	DIC	σ^2_{time}	σ^2_{cohort}	Partition	k	PAR Posterior Mean	95% Lower CI	95% Upper CI
BGLIMM	1	4003	0.79 (0.06)	0.32 (0.11)	None	1	0.99 (0.02)	0.92	0.97
GLAMRE	1	4048	1.97 (0.61)	0.31 (0.13)	Weekday	1	0.46 (0.11)	0.23	0.67
					Weekend	1	0.60 (0.13)	0.36	0.83
	2	**3755**	**7.13 (3.72)**	**0.42 (0.21)**	**Weekday**	**1**	**0.18 (0.07)**	**0.05**	**0.32**
						2	**0.40 (0.09)**	**0.25**	**0.59**
					Weekend	**1**	**0.23 (0.08)**	**0.07**	**0.21**
						2	**0.38 (0.08)**	**0.22**	**0.56**
	3	4058	2.78 (1.78)	0.31 (0.14)	Weekday	1	0.25 (0.13)	0.04	0.54
						2	0.52 (0.14)	0.26	0.82
						3	0.08 (0.19)	-0.26	0.47
					Weekend	1	0.61 (0.29)	0.14	0.99
						2	0.73 (0.19)	0.32	0.93
						3	-0.16 (0.23)	-0.52	0.36

B.B Derivations

B.B.1 Joint Posteriors

Following the prior configuration, we move to flesh out the joint Posterior. The posterior contribution for a given subject i with n_{i} repeat measurements is as follows:

$$\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\theta}_{\text{i}}|\boldsymbol{y}_{\text{i}}) \propto\ & L(\boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\rho}^z, \sigma^2_{\text{time}}, \sigma^2_{\text{cohort}}, \boldsymbol{\gamma}_{\text{time,i}}, \boldsymbol{\gamma}_{\text{cohort}}|\boldsymbol{y}_{\text{i}})\boldsymbol{\pi}(\boldsymbol{\omega}) \\
& \times \boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\sigma^2_{\text{time}})\boldsymbol{\pi}(\sigma^2_{\text{cohort}})\boldsymbol{\pi}(\boldsymbol{\rho}^z)\boldsymbol{\pi}(\boldsymbol{\gamma}_{\text{time,i}}|\boldsymbol{\rho}^z, \sigma^2_{\text{cohort}})\boldsymbol{\pi}(\boldsymbol{\gamma}_{\text{cohort}}|\sigma^2_{\text{cohort}}) \\
& = \left[\prod_{\text{j}=1}^{n_{\text{i}}} \exp\left\{k_{\text{ij}}(\boldsymbol{x}_{\text{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\text{ij}}\boldsymbol{\gamma}_{\text{i}}) - \frac{1}{2}\omega_{\text{ij}}(\boldsymbol{x}_{\text{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\text{ij}}\boldsymbol{\gamma}_{\text{i}})^2\right\} PG(\omega_{\text{ij}}|n_{\text{ij}}, 0)\right] \\
& \times N_p(\boldsymbol{\beta}|\boldsymbol{b}, \boldsymbol{\Sigma}_b) IG(\sigma^2_{\text{time}}|v_0, v_0) IG(\sigma^2_{\text{cohort}}|v_0, v_0) \left(\prod_{k=1}^{K} \text{logistic}\left(\rho^z_k|\mu = 0, s = \frac{1}{2}\right)\right) \\
& \times N_{n_{\text{i}}}\Big(\boldsymbol{\gamma}_{\text{time,i}}|\boldsymbol{0}, \sigma^2_{\text{time}}\boldsymbol{G}\Big) N(\gamma_{\text{cohort,i}}|0, \sigma^2_{\text{cohort}})
\end{aligned}$$

with the following matrix and vector definitions:

$$\begin{aligned}
k_{\text{ij}} &= y_{\text{ij}} - \frac{n_{\text{ij}}}{2}, n_{\text{ij}} = 1 \text{ for binary outcomes} \\
\boldsymbol{z}_{\text{ij}}\boldsymbol{\gamma}_{\text{i}} &= \boldsymbol{z}_{\text{time,j}}\boldsymbol{\gamma}_{\text{time,i}} + \boldsymbol{z}_{\text{cohort,i}}\boldsymbol{\gamma}_{\text{cohort}} \\
\boldsymbol{x}_{\text{ij}} &= \text{ j-th row of matrix } \boldsymbol{X}_{\text{i}} \\
\boldsymbol{z}_{\text{time,j}} &= \text{ j-th row of matrix } \boldsymbol{Z}_{\text{time}} \\
\boldsymbol{z}_{\text{cohort,i}} &= \text{ 1st or any row of matrix } \boldsymbol{Z}_{\text{cohort,i}}
\end{aligned}$$

The joint posterior across all I subjects is given by the following

$$\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) = & \left[\prod_{\text{i}=I}^{I}\prod_{\text{j}=1}^{n_{\text{i}}} \exp\left\{k_{\text{ij}}(\boldsymbol{x}_{\text{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\text{ij}}\boldsymbol{\gamma}_{\text{i}}) - \frac{1}{2}\omega_{\text{ij}}(\boldsymbol{x}_{\text{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\text{ij}}\boldsymbol{\gamma}_{\text{i}})^2\right\} PG(\omega_{\text{ij}}|n_{\text{ij}}, 0)\right] \\
& \times N_p(\boldsymbol{\beta}|\boldsymbol{b}, \boldsymbol{\Sigma}_b) IG(\sigma^2_{\text{time}}|v_0, v_0) IG(\sigma^2_{\text{cohort}}|v_0, v_0) \left(\prod_{k=1}^{K} \text{logistic}\left(\rho_k|\mu = 0, s = \frac{1}{2}\right)\right) \\
& \times \left[\prod_{\text{i}=1}^{I} N_{n_{\text{i}}}\left(\boldsymbol{\gamma}_{\text{time,i}}|\boldsymbol{0}, \sigma^2_{\text{time}}\boldsymbol{G}\right)\right] N_C\left(\boldsymbol{\gamma}_{\text{cohort}}|\boldsymbol{0}, \sigma^2_{\text{cohort}}\mathbb{1}_{C\times C}\right)
\end{aligned}$$

B.B.2 Fixed Effects Posteriors

The posterior of the fixed effects $\boldsymbol{\beta}$ is given by

$$\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\beta}|\boldsymbol{\theta},\boldsymbol{y}) &\propto \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_\mathrm{i}} \exp\left\{k_{\mathrm{ij}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}+\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i}) - \frac{1}{2}\omega_{\mathrm{ij}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}+\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i})^2\right\}\right] N_p(\boldsymbol{\beta}|\boldsymbol{b},\boldsymbol{B}) \\
&\propto \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_\mathrm{i}} \exp\left\{k_{\mathrm{ij}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}+\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i}) - \frac{1}{2}\omega_{\mathrm{ij}}\left((\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta})^2+2(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta})(\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i})\right)\right\}\right] N_p(\boldsymbol{\beta}|\boldsymbol{b},\boldsymbol{B}) \\
&\propto \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_\mathrm{i}} \exp\left\{-\frac{1}{2}\omega_{\mathrm{ij}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta})^2 + (\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta})\left(k_{\mathrm{ij}}-\omega_{\mathrm{ij}}(\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i})\right)\right\}\right] N_p(\boldsymbol{\beta}|\boldsymbol{b},\boldsymbol{B}) \\
&= \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_\mathrm{i}} \exp\left\{-\frac{1}{2}\omega_{\mathrm{ij}}\left((\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta})^2 - 2\frac{k_{\mathrm{ij}}-\omega_{\mathrm{ij}}(\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i})}{\omega_{\mathrm{ij}}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta})\right)\right\}\right] N_p(\boldsymbol{\beta}|\boldsymbol{b},\boldsymbol{B}) \\
&\propto \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_\mathrm{i}} \exp\left\{-\frac{1}{2}\omega_{\mathrm{ij}}\left[\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta} - \left(\frac{k_{\mathrm{ij}}-\omega_{\mathrm{ij}}\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i}}{\omega_{\mathrm{ij}}}\right)\right]^2\right\}\right] N_p(\boldsymbol{\beta}|\boldsymbol{b},\boldsymbol{B})
\end{aligned}$$

Inspired by equation (5) from [103] and page 742 from [4], we will show that for individual i with n_i repeat measurements that:

$$\exp\left\{-\frac{1}{2}(\boldsymbol{l}_\mathrm{i}-\boldsymbol{X}_\mathrm{i}\boldsymbol{\beta})^T\boldsymbol{\Omega}_\mathrm{i}(\boldsymbol{l}_\mathrm{i}-\boldsymbol{X}_\mathrm{i}\boldsymbol{\beta})\right\} = \prod_{\mathrm{j}=1}^{n_\mathrm{i}} \exp\left\{-\frac{1}{2}\omega_{\mathrm{ij}}\left((\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta})^2 - 2\frac{k_{\mathrm{ij}}-\omega_{\mathrm{ij}}(\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_\mathrm{i})}{\omega_{\mathrm{ij}}}(\boldsymbol{x}_{\mathrm{ij}}^T\boldsymbol{\beta})\right)\right\}$$

$$\text{where } \boldsymbol{l}_\mathrm{i} = \begin{bmatrix} \frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}} - \boldsymbol{z}_{\mathrm{i}1}\boldsymbol{\gamma}_\mathrm{i} \\ \vdots \\ \frac{k_{\mathrm{i}J}}{\omega_{\mathrm{i}J}} - \boldsymbol{z}_{\mathrm{i}J}\boldsymbol{\gamma}_\mathrm{i} \end{bmatrix}$$

$$\boldsymbol{\Omega}_\mathrm{i} = \begin{bmatrix} \omega_{\mathrm{i}1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \omega_{\mathrm{i}n_\mathrm{i}} \end{bmatrix} = diag(\omega_{\mathrm{i}1},\dots,\omega_{\mathrm{i}n_\mathrm{i}})$$

We begin with the left hand side of the expression:

$$
\begin{aligned}
&\exp\left\{-\frac{1}{2}(\boldsymbol{l}_{\mathrm{i}}-\boldsymbol{X}_{\mathrm{i}}\boldsymbol{\beta})^{T}\boldsymbol{\Omega}_{\mathrm{i}}(\boldsymbol{l}_{\mathrm{i}}-\boldsymbol{X}_{\mathrm{i}}\boldsymbol{\beta})\right\} \\
&\quad=\exp\left\{-\frac{1}{2}\begin{bmatrix}\frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}}-\boldsymbol{z}_{\mathrm{i}1}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}1}\boldsymbol{\beta}\\ \cdots \\ \frac{k_{\mathrm{i}J}}{\omega_{\mathrm{i}J}}-\boldsymbol{z}_{\mathrm{i}J}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}J}\boldsymbol{\beta}\end{bmatrix}^{T}\begin{bmatrix}\omega_{\mathrm{i}1} & \ldots & 0\\ \vdots & \ddots & \vdots\\ 0 & \ldots & \omega_{\mathrm{i}J}\end{bmatrix}\begin{bmatrix}\frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}}-\boldsymbol{z}_{\mathrm{i}1}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}1}\boldsymbol{\beta}\\ \cdots \\ \frac{k_{\mathrm{i}J}}{\omega_{\mathrm{i}J}}-\boldsymbol{z}_{\mathrm{i}J}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}J}\boldsymbol{\beta}\end{bmatrix}\right\} \\
&\quad=\exp\left\{-\frac{1}{2}\begin{bmatrix}\omega_{\mathrm{i}1}\left(\frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}}-\boldsymbol{z}_{\mathrm{i}1}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}1}\boldsymbol{\beta}\right)\\ \cdots \\ \omega_{\mathrm{i}J}\left(\frac{k_{\mathrm{i}J}}{\omega_{\mathrm{i}J}}-\boldsymbol{z}_{\mathrm{i}J}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}J}\boldsymbol{\beta}\right)\end{bmatrix}^{T}\begin{bmatrix}\frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}}-\boldsymbol{z}_{\mathrm{i}1}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}1}\boldsymbol{\beta}\\ \cdots \\ \frac{k_{\mathrm{i}J}}{\omega_{\mathrm{i}J}}-\boldsymbol{z}_{\mathrm{i}J}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{i}J}\boldsymbol{\beta}\end{bmatrix}\right\} \\
&\quad=\exp\left\{-\frac{1}{2}\sum_{\mathrm{j}=1}^{n_{\mathrm{i}}}\omega_{\mathrm{ij}}\left(\frac{k_{\mathrm{ij}}}{\omega_{\mathrm{ij}}}-\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_{\mathrm{i}}-\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}\right)^{2}\right\} \\
&\quad=\prod_{\mathrm{j}=1}^{n_{\mathrm{i}}}\exp\left\{-\frac{1}{2}\omega_{\mathrm{ij}}\left[\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}-\left(\frac{k_{\mathrm{ij}}-\omega_{\mathrm{ij}}\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_{\mathrm{i}}}{\omega_{\mathrm{ij}}}\right)\right]^{2}\right\}
\end{aligned}
$$

From above, we have just shown the equivalence of the LHS expression and RHS expression. Formally, in matrix notation, we have the following:

$$
\begin{aligned}
\boldsymbol{\pi}(\boldsymbol{\beta}|\boldsymbol{\theta},\boldsymbol{y}) &\propto \exp\left\{-\frac{1}{2}\sum_{\mathrm{i}=1}^{I}(\boldsymbol{l}_{\mathrm{i}}-\boldsymbol{X}_{\mathrm{i}}\boldsymbol{\beta})^{T}\boldsymbol{\Omega}_{\mathrm{i}}(\boldsymbol{l}_{\mathrm{i}}-\boldsymbol{X}_{\mathrm{i}}\boldsymbol{\beta})\right\}N_{p}(\boldsymbol{\beta}|\boldsymbol{b},\boldsymbol{\Sigma}_{b}) \\
\text{where } \boldsymbol{l}_{\mathrm{i}} &= \begin{bmatrix}\frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}}-\boldsymbol{z}_{\mathrm{i}1}\boldsymbol{\gamma}_{\mathrm{i}}\\ \vdots \\ \frac{k_{\mathrm{i}J}}{\omega_{\mathrm{i}J}}-\boldsymbol{z}_{\mathrm{i}J}\boldsymbol{\gamma}_{\mathrm{i}}\end{bmatrix} \\
\boldsymbol{\Omega}_{\mathrm{i}} &= diag(\omega_{\mathrm{i}1},\ldots,\omega_{\mathrm{i}J})
\end{aligned}
$$

Based on page (4) of [104], we should have

$$
\begin{aligned}
\boldsymbol{\beta}|\boldsymbol{\theta},\boldsymbol{y} &\sim N_{p}\left(\boldsymbol{\mu}_{\beta},\boldsymbol{\Sigma}_{\beta}\right) \\
\boldsymbol{\Sigma}_{\beta} &= \left(\boldsymbol{\Sigma}_{b}^{-1}+\sum_{\mathrm{i}=1}^{N}\boldsymbol{X}_{\mathrm{i}}^{T}\boldsymbol{\Omega}_{\mathrm{i}}\boldsymbol{X}_{\mathrm{i}}\right)^{-1} \\
\boldsymbol{\mu}_{\beta} &= \boldsymbol{\Sigma}_{\beta}\left(\boldsymbol{\Sigma}_{b}^{-1}\boldsymbol{b}+\sum_{\mathrm{i}=1}^{N}\boldsymbol{X}_{\mathrm{i}}^{T}\boldsymbol{\Omega}_{\mathrm{i}}\boldsymbol{l}_{\mathrm{i}}\right)
\end{aligned}
$$

B.B.3 Time Random Effect Posteriors

For convenience, we define the following

$$\begin{aligned}
\tilde{\boldsymbol{z}}_{\mathrm{ij}}\tilde{\boldsymbol{\gamma}}_{\mathrm{i}} &= \boldsymbol{z}_{\mathrm{time,j}}\boldsymbol{\gamma}_{\mathrm{time,i}} \\
\tilde{\boldsymbol{x}}_{\mathrm{ij}}\tilde{\boldsymbol{\beta}} &= \boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\mathrm{cohort,i}}\boldsymbol{\gamma}_{\mathrm{cohort}} \\
\Rightarrow \boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\mathrm{time,j}}\boldsymbol{\gamma}_{\mathrm{time,i}} + \boldsymbol{z}_{\mathrm{cohort,i}}\boldsymbol{\gamma}_{\mathrm{cohort}} &= \tilde{\boldsymbol{z}}_{\mathrm{ij}}\tilde{\boldsymbol{\gamma}}_{\mathrm{i}} + \tilde{\boldsymbol{x}}_{\mathrm{ij}}\tilde{\boldsymbol{\beta}}
\end{aligned}$$

To derive the posterior of the random effects $\boldsymbol{\gamma}_{\mathrm{time,i}}$, we proceed similarly as we had when deriving the posterior of $\boldsymbol{\beta}$:

$$\begin{aligned}
&\boldsymbol{\pi}(\boldsymbol{\gamma}_{\mathrm{time,i}}|\boldsymbol{\theta}, \boldsymbol{y}_{\mathrm{i}}) \\
&\quad\propto \left[\prod_{\mathrm{j}=1}^{n_{\mathrm{i}}} \exp\left\{k_{\mathrm{ij}}\left(\tilde{\boldsymbol{z}}_{\mathrm{ij}}\tilde{\boldsymbol{\gamma}}_{\mathrm{i}} + \tilde{\boldsymbol{x}}_{\mathrm{ij}}\tilde{\boldsymbol{\beta}}\right) - \frac{1}{2}\omega_{\mathrm{ij}}\left(\tilde{\boldsymbol{z}}_{\mathrm{ij}}\tilde{\boldsymbol{\gamma}}_{\mathrm{i}} + \tilde{\boldsymbol{x}}_{\mathrm{ij}}\tilde{\boldsymbol{\beta}}\right)^2\right\}\right] N_{n_{\mathrm{i}}}(\boldsymbol{\gamma}_{\mathrm{time,i}}|\boldsymbol{0}, \sigma^2_{\mathrm{time}}\boldsymbol{G}) \\
&\quad\propto \exp\left\{-\frac{1}{2}\left(\tilde{\boldsymbol{l}}_{\mathrm{i}} - \boldsymbol{Z}_{\mathrm{time,i}}\boldsymbol{\gamma}_{\mathrm{time,i}}\right)^T \boldsymbol{\Omega}_{\mathrm{i}}\left(\tilde{\boldsymbol{l}}_{\mathrm{i}} - \boldsymbol{Z}_{\mathrm{time,i}}\boldsymbol{\gamma}_{\mathrm{time,i}}\right)\right\} N_{n_{\mathrm{i}}}(\boldsymbol{\gamma}_{\mathrm{time,i}}|\boldsymbol{0}, \sigma^2_{\mathrm{time}}\boldsymbol{G}) \\
&\text{where } \tilde{\boldsymbol{l}}_{\mathrm{i}} = \begin{bmatrix} \frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}} - \tilde{\boldsymbol{x}}_{\mathrm{i}1}\tilde{\boldsymbol{\beta}} \\ \vdots \\ \frac{k_{\mathrm{i}n_{\mathrm{i}}}}{\omega_{\mathrm{i}n_{\mathrm{i}}}} - \tilde{\boldsymbol{x}}_{\mathrm{i}J}\tilde{\boldsymbol{\beta}} \end{bmatrix} \\
&\boldsymbol{\Omega}_{\mathrm{i}} = diag(\omega_{\mathrm{i}1}, \ldots, \omega_{\mathrm{i}J}) \\
&k_{\mathrm{ij}} = y_{\mathrm{ij}} - \frac{1}{2}n_{\mathrm{ij}}, \text{ where } n_{\mathrm{ij}} = 1 \text{ for bernoulli trials}
\end{aligned}$$

And as before, we have

$$\begin{aligned}
\boldsymbol{\gamma}_{\mathrm{time,i}}|\boldsymbol{\theta}, \boldsymbol{y}_{\mathrm{i}} &\sim N_{n_{\mathrm{i}}}(\boldsymbol{\mu}_{\gamma_{\mathrm{time,i}}}, \boldsymbol{\Sigma}_{\gamma_{\mathrm{time,i}}}) \\
\boldsymbol{\Sigma}_{\gamma_{\mathrm{time,i}}} &= \left((\sigma^2_{\mathrm{time}}\boldsymbol{G})^{-1} + \boldsymbol{Z}^T_{\mathrm{time,i}}\Omega_{\mathrm{i}}\boldsymbol{Z}_{\mathrm{time,i}}\right)^{-1} \\
\boldsymbol{\mu}_{\gamma_{\mathrm{time,i}}} &= \boldsymbol{\Sigma}_{\gamma_{\mathrm{time,i}}}\left((\sigma^2_{\mathrm{time}}\boldsymbol{G})^{-1}\boldsymbol{0} + \boldsymbol{Z}^T_{\mathrm{time,i}}\Omega_{\mathrm{i}}\tilde{\boldsymbol{l}}_{\mathrm{i}}\right) \\
&= \boldsymbol{\Sigma}_{\gamma_{\mathrm{time,i}}}\left(\boldsymbol{Z}^T_{\mathrm{time,i}}\Omega_{\mathrm{i}}\tilde{\boldsymbol{l}}_{\mathrm{i}}\right)
\end{aligned}$$

B.B.4 Cohort Random Effect Posteriors

For convenience, we define the following

$$\begin{aligned}
\tilde{\boldsymbol{z}}_{\mathrm{ij}}\tilde{\boldsymbol{\gamma}}_{\mathrm{i}} &= \boldsymbol{z}_{\mathrm{time,j}}\boldsymbol{\gamma}_{\mathrm{time,i}} \\
\tilde{\boldsymbol{x}}_{\mathrm{ij}}\tilde{\boldsymbol{\beta}} &= \boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\mathrm{cohort,i}}\boldsymbol{\gamma}_{\mathrm{cohort}} \\
\Rightarrow \boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta} + \boldsymbol{z}_{\mathrm{time,j}}\boldsymbol{\gamma}_{\mathrm{time,i}} + \boldsymbol{z}_{\mathrm{cohort,i}}\boldsymbol{\gamma}_{\mathrm{cohort}} &= \tilde{\boldsymbol{z}}_{\mathrm{ij}}\tilde{\boldsymbol{\gamma}}_{\mathrm{i}} + \tilde{\boldsymbol{x}}_{\mathrm{ij}}\tilde{\boldsymbol{\beta}}
\end{aligned}$$

To derive the posterior of the random effects $\boldsymbol{\gamma}_{\mathrm{cohort}}$, we proceed similarly as we had when deriving the posterior of $\boldsymbol{\beta}$:

$$\begin{aligned}
&\boldsymbol{\pi}(\boldsymbol{\gamma}_{\mathrm{cohort}}|\boldsymbol{\theta},\boldsymbol{y}) \\
&\quad\propto \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_{\mathrm{i}}}\exp\left\{k_{\mathrm{ij}}\left(\breve{\boldsymbol{z}}_{\mathrm{ij}}\breve{\boldsymbol{\gamma}}_{\mathrm{i}} + \breve{\boldsymbol{x}}_{\mathrm{ij}}\breve{\boldsymbol{\beta}}\right) - \frac{1}{2}\omega_{\mathrm{ij}}\left(\breve{\boldsymbol{z}}_{\mathrm{ij}}\breve{\boldsymbol{\gamma}}_{\mathrm{i}} + \breve{\boldsymbol{x}}_{\mathrm{ij}}\breve{\boldsymbol{\beta}}\right)^2\right\}\right] N_C\left(\boldsymbol{\gamma}_{\mathrm{cohort}}|\boldsymbol{0},\sigma^2_{\mathrm{cohort}}\mathbb{1}_{C\times C}\right) \\
&\quad\propto \exp\left\{-\frac{1}{2}\sum_{\mathrm{i}=1}^{I}\left(\breve{\boldsymbol{l}}_{\mathrm{i}} - \boldsymbol{Z}_{\mathrm{cohort,i}}\boldsymbol{\gamma}_{\mathrm{cohort}}\right)^T\boldsymbol{\Omega}_{\mathrm{i}}\left(\breve{\boldsymbol{l}}_{\mathrm{i}} - \boldsymbol{Z}_{\mathrm{cohort,i}}\boldsymbol{\gamma}_{\mathrm{cohort}}\right)\right\} N_C\left(\boldsymbol{\gamma}_{\mathrm{cohort}}|\boldsymbol{0},\sigma^2_{\mathrm{cohort}}\mathbb{1}_{C\times C}\right)
\end{aligned}$$

where $\breve{\boldsymbol{l}}_{\mathrm{i}} = \begin{bmatrix} \frac{k_{\mathrm{i}1}}{\omega_{\mathrm{i}1}} - \breve{\boldsymbol{x}}_{\mathrm{i}1}\breve{\boldsymbol{\beta}} \\ \vdots \\ \frac{k_{\mathrm{i},J}}{\omega_{\mathrm{i},J}} - \breve{\boldsymbol{x}}_{\mathrm{i},J}\breve{\boldsymbol{\beta}} \end{bmatrix}$

$\boldsymbol{\Omega}_{\mathrm{i}} = diag(\omega_{\mathrm{i}1},\ldots,\omega_{\mathrm{i}J})$

$k_{\mathrm{ij}} = y_{\mathrm{ij}} - \frac{1}{2}n_{\mathrm{ij}}$, where $n_{\mathrm{ij}} = 1$ for bernoulli trials

And as before, we have

$$\begin{aligned}
\boldsymbol{\gamma}_{\mathrm{cohort}}|\boldsymbol{\theta},\boldsymbol{y} &\sim N_C\left(\boldsymbol{\mu}_{\mathrm{cohort}},\boldsymbol{\Sigma}_{\mathrm{cohort}}\right) \\
\boldsymbol{\Sigma}_{\mathrm{cohort}} &= \left(\frac{1}{\sigma^2_{\mathrm{cohort}}}\mathbb{1}_{C\times C} + \sum_{\mathrm{i}=1}^{I}\boldsymbol{Z}^T_{\mathrm{cohort,i}}\boldsymbol{\Omega}_{\mathrm{i}}\boldsymbol{Z}_{\mathrm{cohort,i}}\right)^{-1} \\
\boldsymbol{\mu}_{\gamma_{\mathrm{cohort}}} &= \boldsymbol{\Sigma}_{\mathrm{cohort}}\left(\frac{1}{\sigma^2_{\mathrm{cohort}}}\mathbb{1}_{C\times C}\boldsymbol{0} + \sum_{\mathrm{i}=1}^{I}\boldsymbol{Z}^T_{\mathrm{cohort,i}}\boldsymbol{\Omega}_{\mathrm{i}}\breve{\boldsymbol{l}}_{\mathrm{i}}\right) \\
&= \boldsymbol{\Sigma}_{\mathrm{cohort}}\left(\sum_{\mathrm{i}=1}^{I}\boldsymbol{Z}^T_{\mathrm{cohort,i}}\boldsymbol{\Omega}_{\mathrm{i}}\breve{\boldsymbol{l}}_{\mathrm{i}}\right)
\end{aligned}$$

B.B.5 Polya-Gamma Latent Posteriors

The posterior of $\boldsymbol{\omega}$ is given by

$$\begin{aligned}\boldsymbol{\pi}(\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{y}) &= \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_{\mathrm{i}}}\exp\left\{k_{\mathrm{ij}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}+\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_{\mathrm{i}})-\frac{1}{2}\omega_{\mathrm{ij}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}+\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_{\mathrm{i}})^2\right\}PG(\boldsymbol{w}_{\mathrm{ij}}|n_{\mathrm{ij}},0)\right] \\ &\propto \left[\prod_{\mathrm{i}=1}^{I}\prod_{\mathrm{j}=1}^{n_{\mathrm{i}}}\exp\left\{-\frac{1}{2}\omega_{\mathrm{ij}}(\boldsymbol{x}_{\mathrm{ij}}\boldsymbol{\beta}+\boldsymbol{z}_{\mathrm{ij}}\boldsymbol{\gamma}_{\mathrm{i}})^2\right\}PG(\omega_{\mathrm{ij}}|n_{\mathrm{ij}},0)\right]\end{aligned}$$

Based on [68], we get the conjugate posterior

$$\omega_{\mathrm{ij}}|\boldsymbol{\theta},\boldsymbol{y}\sim PG(n_{\mathrm{ij}}=1,\boldsymbol{x}_{\mathrm{ij}}^T\boldsymbol{\beta}+\boldsymbol{z}_{\mathrm{ij}}^T\boldsymbol{\gamma}_{\mathrm{i}})$$

B.B.6 Partial Autocorrelation Posteriors

The posterior of $\boldsymbol{\rho}^z$ is given by

$$\boldsymbol{\pi}(\boldsymbol{\rho}^z|\boldsymbol{\theta},\boldsymbol{y})\propto\left[\prod_{\mathrm{i}=1}^{I}N_{n_{\mathrm{i}}}\left(\boldsymbol{\gamma}_{\mathrm{time,i}}|\boldsymbol{0},\sigma^2_{\mathrm{time}}\boldsymbol{G}\right)\right]\left[\prod_{k=1}^{K}\mathrm{logistic}\left(\rho_k|\mu=0,s=\frac{1}{2}\right)\right]$$
$$\boldsymbol{G}=r(\tanh(\boldsymbol{\rho}_k^z)),r(.)=\text{ recurrence relation from [30]}$$

We will sample from $\boldsymbol{\rho}^z|\boldsymbol{\theta},\boldsymbol{y}$ via Metropolis Hastings using the configuration below:

$$\begin{aligned}\text{Target: }&\boldsymbol{\pi}(\boldsymbol{\rho}^z|\boldsymbol{\theta},\boldsymbol{y})\\ \text{Proposal: }&N_K\left(\boldsymbol{\rho}^{z^{(t)}}|\boldsymbol{\rho}^{z^{(t-1)}},\sigma^2_{\mathrm{prop}}\boldsymbol{I}_K\right)\end{aligned}$$

Acceptance prob:

$$\begin{aligned}\alpha\left(\boldsymbol{\rho}^{z^{(t-1)}},\boldsymbol{\rho}^{z^{(t)}}\right)&=\frac{\boldsymbol{\pi}(\boldsymbol{\rho}^{z^{(t)}}|\boldsymbol{\theta},\boldsymbol{y})}{\boldsymbol{\pi}(\boldsymbol{\rho}^{z^{(t-1)}}|\boldsymbol{\theta},\boldsymbol{y})}\\ &=\frac{\left[\prod_{\mathrm{i}=1}^{I}N_{n_{\mathrm{i}}}\left(\boldsymbol{\gamma}_{\mathrm{time,i}}|\boldsymbol{0},\sigma^2_{\mathrm{time}}\boldsymbol{G}^{(t)}\right)\right]\left[\prod_{k=1}^{K}\mathrm{logistic}\left(\rho_k^{z,(t)}|\mu=0,s=\frac{1}{2}\right)\right]}{\left[\prod_{\mathrm{i}=1}^{I}N_{n_{\mathrm{i}}}\left(\boldsymbol{\gamma}_{\mathrm{time,i}}|\boldsymbol{0},\sigma^2_{\mathrm{time}}\boldsymbol{G}^{(t-1)}\right)\right]\left[\prod_{k=1}^{K}\mathrm{logistic}\left(\rho_k^{z,(t-1)}|\mu=0,s=\frac{1}{2}\right)\right]}\end{aligned}$$

B.B.7 Variance From Time Posterior

With conjugate priors, we have the following posterior for σ^2_{time}

$$\begin{aligned}
\boldsymbol{\pi}(\sigma^2_{\text{time}}|\boldsymbol{\theta},\boldsymbol{y}) &\propto \left[\prod_{\text{i}=1}^{I} N_{n_{\text{i}}}\left(\boldsymbol{\gamma}_{\text{time,i}}|\mathbf{0},\sigma^2_{\text{time}}\boldsymbol{G}\right)\right] IG(\sigma^2_{\text{time}}|v_0,v_0) \\
&\propto \left[\prod_{\text{i}=1}^{I}(\det(\sigma^2_{\text{time}}\boldsymbol{G}))^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2_{\text{time}}}\boldsymbol{\gamma}^T_{\text{time,i}}\boldsymbol{G}^{-1}\boldsymbol{\gamma}_{\text{time,i}}\right\}\right](\sigma^2_{\text{time}})^{-(v_0+1)}\exp\left\{-\frac{v_0}{\sigma^2_{\text{time}}}\right\} \\
&\propto \frac{1}{(\sigma^2_{\text{time}})^{\frac{IJ}{2}}}\exp\left\{-\frac{1}{2\sigma^2_{\text{time}}}\sum_{\text{i}=1}^{I}\boldsymbol{\gamma}^T_{\text{time,i}}\boldsymbol{G}^{-1}\boldsymbol{\gamma}_{\text{time,i}}\right\}(\sigma^2_{\text{time}})^{-(v_0+1)}\exp\left\{-\frac{v_0}{\sigma^2_{\text{time}}}\right\} \\
&= (\sigma^2_{\text{time}})^{-\left(\frac{IJ}{2}+v_0\right)-1}\exp\left\{-\frac{1}{\sigma^2_{\text{time}}}\frac{2v_0+\sum_{\text{i}=1}^{I}\boldsymbol{\gamma}^T_{\text{time,i}}\boldsymbol{G}^{-1}\boldsymbol{\gamma}_{\text{time,i}}}{2}\right\} \\
\Rightarrow \sigma^2_{\text{time}}|\boldsymbol{\theta},\boldsymbol{y} &\sim IG\left(\left(\frac{IJ}{2}+v_0\right),\frac{2v_0+\sum_{\text{i}=1}^{I}\boldsymbol{\gamma}^T_{\text{time,i}}\boldsymbol{G}^{-1}\boldsymbol{\gamma}_{\text{time,i}}}{2}\right)
\end{aligned}$$

B.B.8 Variance From Cohort Posterior

Similar to σ^2_{time}, we have the following posterior for σ^2_{cohort}

$$\begin{aligned}
\boldsymbol{\pi}(\sigma^2_{\gamma_{\text{cohort}}}|\boldsymbol{\theta},\boldsymbol{y}) &\propto N_C\left(\boldsymbol{\gamma}_{\text{cohort}}|\mathbf{0},\sigma^2_{\text{cohort}}\mathbb{1}_{C\times C}\right)IG(\sigma^2_{\text{cohort}}|v_0,v_0) \\
&\propto (\det(\sigma^2_{\text{cohort}}\mathbb{1}_{C\times C}))^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2_{\text{cohort}}}\boldsymbol{\gamma}^T_{\text{cohort}}\mathbb{1}^{-1}_{C\times C}\boldsymbol{\gamma}_{\text{cohort}}\right\}(\sigma^2_{\text{cohort}})^{-(v_0+1)}\exp\left\{-\frac{v_0}{\sigma^2_{\text{cohort}}}\right\} \\
&= (\sigma^2_{\text{cohort}})^{-(\frac{C}{2}+v_0)-1}\exp\left\{-\frac{1}{\sigma^2_{\text{cohort}}}\frac{2v_0+\boldsymbol{\gamma}^T_{\text{cohort}}\mathbb{1}^{-1}_{C\times C}\boldsymbol{\gamma}_{\text{cohort}}}{2}\right\} \\
\Rightarrow \sigma^2_{\gamma_{\text{cohort}}}|\boldsymbol{\theta},\boldsymbol{y} &\sim IG\left(\left(\frac{C}{2}+v_0\right),\frac{2v_0+\boldsymbol{\gamma}^T_{\text{cohort}}\mathbb{1}^{-1}_{C\times C}\boldsymbol{\gamma}_{\text{cohort}}}{2}\right)
\end{aligned}$$

B.C SAS Script

```
proc bglimm data=final nmc=3000 seed=901214 Statistics=sum  Stats=int;
class y treatment subject cohort time;
model y(event = "1") = treatment x_0 / dist=binary link=logit;
random time/ type=ar(1) subject=subject nuts;
random intercept/ subject=cohort nuts;
run;
```